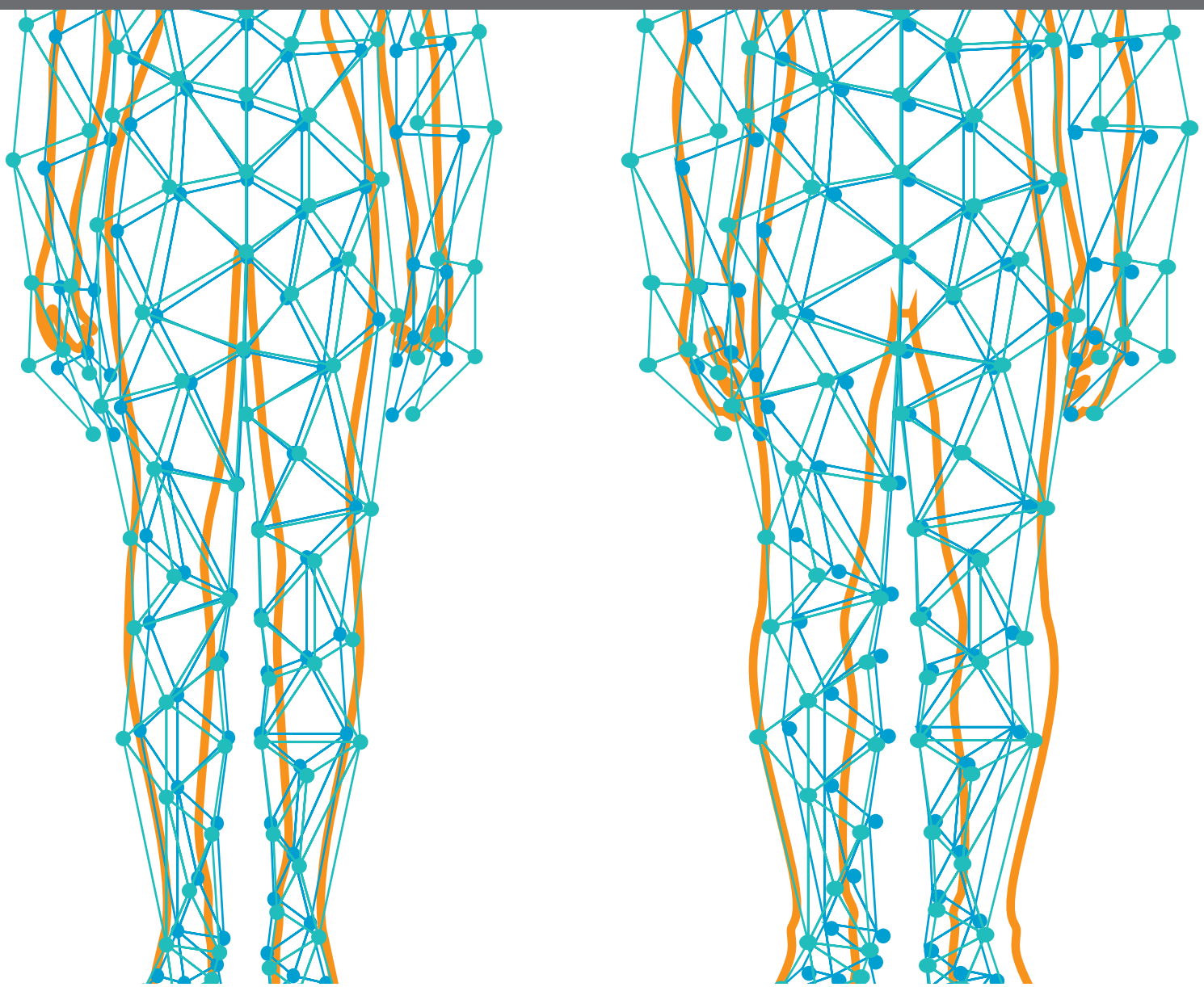
CLINICAL APPLICATION OF ARTIFICIAL INTELLIGENCE IN EMERGENCY AND CRITICAL CARE MEDICINE, VOLUME II

EDITED BY: Zhongheng Zhang, Rahul Kashyap, Nan Liu, Longxiang Su and Qinghe Meng

PUBLISHED IN: Frontiers in Medicine



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-261-3
DOI 10.3389/978-2-88976-261-3

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

CLINICAL APPLICATION OF ARTIFICIAL INTELLIGENCE IN EMERGENCY AND CRITICAL CARE MEDICINE, VOLUME II

Topic Editors:
Zhongheng Zhang, Sir Run Run Shaw Hospital, China
Rahul Kashyap, Mayo Clinic, United States
Nan Liu, National University of Singapore, Singapore
Longxiang Su, Peking Union Medical College Hospital (CAMS), China
Qinghe Meng, Upstate Medical University, United States

Citation: Zhang, Z., Kashyap, R., Liu, N., Su, L., Meng, Q., (2022). Clinical Application of Artificial Intelligence in Emergency and Critical Care Medicine, Volume II. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-261-3

Table of Contents

frontiers | Frontiers in Medicine

EDITORIAL
published: 06 May 2022
doi: 10.3389/fmed.2022.910163

Editorial: Clinical Application of Artificial Intelligence in Emergency and Critical Care Medicine, Volume II

Zhongheng Zhang[1]*, *Rahul Kashyap*[2,3], *Nan Liu*[4], *Longxiang Su*[5] and *Qinghe Meng*[6]

[1] Department of Emergency Medicine, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [2] Critical Care Independent Multidisciplinary Program, Mayo Clinic, Rochester, MN, United States, [3] Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, United States, [4] Programme in Health Services and Systems Research, Duke-National University of Singapore Medical School, Singapore, Singapore, [5] State Key Laboratory of Complex Severe and Rare Diseases, Department of Critical Care Medicine, Peking Union Medical College Hospital, Chinese Academy of Medical Science, Peking Union Medical College, Beijing, China, [6] Department of Surgery, SUNY Upstate Medical University, Syracuse, NY, United States

Keywords: artificial intelligence, critical care, emergency and critical care, heterogeneity, prediction model

Editorial on the Research Topic

Clinical Application of Artificial Intelligence in Emergency and Critical Care Medicine, Volume II

We are excited to witness the successful publication of the second volume of the topic issue entitled *Clinical Application of Artificial Intelligence in Emergency and Critical Care Medicine*. Artificial intelligence (AI) continues to be a hot spot in emergency and critical care medicine, which provides numerous tools and algorithms to learn from data (1, 2). There are three major categories of machine learning (ML) algorithms including supervised learning (3), unsupervised learning, and reinforcement learning (4). These three main categories of ML methods can be utilized to solve clinical questions that cannot be well captured by human intuition.

Supervised learning is defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately, which occurs as part of the cross-validation process (5). The labels in the critical care setting include mortality, the occurrence of postoperative acute kidney injury, delirium, and the development of CRBSI (6–8). In this topic issue, several studies utilized a supervised ML algorithm to predict interested clinical outcomes. Cheng et al. developed a deep learning algorithm (ResNet50-V2) for the detection of abdominal free fluid in Morison's pouch during focused assessment with sonography in trauma. While the model is not externally validated, such seminal attempts will trigger further explorations on the use of radiomics in a critical care setting. Li et al. developed an ML algorithm using eXtreme Gradient Boosting (XGBoost) to predict acute renal failure after acute aortic syndrome surgery. The authors further demonstrated that the XGBoost algorithm was superior to the conventional logistic regression models.

Unlike specialty areas where patients are usually homogenous and thus can benefit from evidence-based medicine (EBM). Under the paradigm of EBM, a group of homogenous population are studied as a whole, and the effects drawn from the population are assumed to be equal in each individual subject (9). In other words, the individual effects are equal to the population mean. In such an ideal situation, a large well-conducted randomized controlled trial can provide the most effective treatment options for the target population. However, the individual effects deviate from the population mean with increasing heterogeneity of the study population. This is what happens in the critical care setting. Although critically ill patients are usually presenting similar clinical manifestations which we group as syndromes such as sepsis and acute respiratory distress syndrome (ARDS), these syndromes usually display significant heterogeneity in their clinical presentations and responses to treatment (10, 11). By a *post-hoc*

Edited and reviewed by:
Ata Murat Kaynar,
University of Pittsburgh, United States

***Correspondence:**
Zhongheng Zhang
zh_zhang1984@zju.edu.cn

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *01 April 2022*
Accepted: *08 April 2022*
Published: *06 May 2022*

Citation:
Zhang Z, Kashyap R, Liu N, Su L and Meng Q (2022) Editorial: Clinical Application of Artificial Intelligence in Emergency and Critical Care Medicine, Volume II. Front. Med. 9:910163. doi: 10.3389/fmed.2022.910163

analysis of a large ARDS trial, Xing et al. identified four phenotypes of ARDS, and the "hyperinflammatory Anasarca" phenotype was found to benefit from a conservative fluid management strategy. The strength of the study was the utility of a randomized trial. The randomization procedure ensured the exchangeability of the study population between treated and untreated groups. Latent class analysis was performed on an ARDS population, which identified three subphenotypes. A subtype characterized by worse renal function and higher central venous pressure (CVP) was found to be beneficial from diuretic use. Collectively, these studies demonstrated the successful application of the ML algorithm to identify subtypes of a heterogeneous population. Interestingly, these subtypes show both prognostic and predictive enrichment for clinical interventions.

The above-mentioned ML algorithms can only be helpful in risk stratification. We know a patient is at risk of death, but do not know how to take action to minimize the risks. In this regard, the third important algorithm called reinforcement learning comes to solve this clinical question (12, 13). Reinforcement learning is a type of machine learning technique where a computer agent learns to perform a task through repeated trial and error interactions with a dynamic environment. This learning approach enables the agent to make a series of decisions that maximize a reward metric for the task without human intervention and without being explicitly programmed to achieve the task. Fluid treatment is an important measure to improve the survival outcome of sepsis. However, the major challenge in fluid treatment is that it requires sequential decision rules to be made by considering the changing status of the subject, as well as its response to previous fluid challenges. Thus, there is no one-size-fits-all model for fluid therapy for sepsis patients. In this regard, Su et al. developed a reinforcement learning algorithm to optimize fluid treatment strategy. They further demonstrated that patients who received the fluid volume close to the estimated volume had the lowest mortality rate. However, since the algorithm is not tested in prospective studies, its clinical utility is still limited.

In conclusion, the topic issues on *Clinical Application of Artificial Intelligence in Emergency and Critical Care Medicine* witnessed tremendous success in the previous two volumes, suggesting continuous interest in this topic in critical care society. Volume III of the topic issue is still open for submission. With the developments in AI technology, it will change the practice pattern of emergency and critical care in the future.

AUTHOR CONTRIBUTIONS

ZZ conceived the idea and drafted the manuscript. NL, RK, and QM critically reviewed the paper and interpretation of the work. LS provided insightful comments for the work. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Lal A, Pinevich Y, Gajic O, Herasevich V, Pickering B. Artificial intelligence and computer simulation models in critical illness. *World J Crit Care Med.* (2020) 9:13–9. doi: 10.5492/wjccm.v9.i2.13
2. Hashimoto DA, Witkowski E, Gao L, Meireles O, Rosman G. Artificial intelligence in anesthesiology: current techniques, clinical applications, and limitations. *Anesthesiology.* (2020) 132:379–94. doi: 10.1097/ALN.0000000000002960
3. Zhang Z, Chen L, Xu P, Hong Y. Predictive analytics with ensemble modeling in laparoscopic surgery: a technical note. *Laparosc Endosc Rob Surgery.* (2022) 5:25–34. doi: 10.1016/j.lers.2021.12.003
4. Zhang Z, Navarese EP, Zheng B, Meng Q, Liu N, Ge H, et al. Analytics with artificial intelligence to advance the treatment of acute respiratory distress syndrome. *J Evid Based Med.* (2020) 13:301–12. doi: 10.1111/jebm.12418
5. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol.* (2019) 6:2374289519873088. doi: 10.1177/2374289519873088
6. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A machine learning algorithm to predict severe sepsis and septic shock: development, implementation, and impact on clinical practice. *Crit Care Med.* (2019) 47:1485–92. doi: 10.1097/CCM.0000000000003891
7. Garnica O, Gómez D, Ramos V, Hidalgo JI, Ruiz-Giardín JM. Diagnosing hospital bacteraemia in the framework of predictive, preventive and personalised medicine using electronic health records and machine learning classifiers. *EPMA J.* (2021) 12:1–17. doi: 10.1007/s13167-021-00252-3
8. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care.* (2019) 23:112. doi: 10.1186/s13054-019-2411-z
9. Vincent J-L, Singer M, Einav S, Moreno R, Wendon J, Teboul J-L, et al. Equilibrating SSC guidelines with individualized care. *Crit Care.* (2021) 25:397. doi: 10.1186/s13054-021-03813-0
10. Vignon P, Evrard B, Asfar P, Busana M, Calfee CS, Coppola S, et al. Fluid administration and monitoring in ARDS: which management? *Intensive Care Med.* (2020) 46:2252–64. doi: 10.1007/s00134-020-06310-0
11. Bhavani SV, Wolfe KS, Hrusch CL, Greenberg JA, Krishack PA, Lin J, et al. Temperature trajectory subphenotypes correlate with immune responses in patients with sepsis. *Crit Care Med.* (2020) 48:1645–53. doi: 10.1097/CCM.0000000000004610
12. Lu M, Shahn Z, Sow D, Doshi-Velez F, Lehman L-WH. Is deep reinforcement learning ready for practical applications in healthcare? A sensitivity analysis of duel-DDQN for hemodynamic management in sepsis patients. *AMIA Annu Symp Proc.* (2020) 2020:773–82. doi: 10.48550/arXiv.2005.04301
13. Peng X, Ding Y, Wihl D, Gottesman O, Komorowski M, Lehman L-WH, et al. Improving sepsis treatment strategies by combining deep and kernel-based reinforcement learning. *AMIA Annu Symp Proc.* (2018) 2018:887–96. doi: 10.48550/arXiv.1901.04670

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 25 August 2021
doi: 10.3389/fmed.2021.727910

ARDS Patients Exhibiting a "Hyperinflammatory Anasarca" Phenotype Could Benefit From a Conservative Fluid Management Strategy

Chun-yan Xing[1,2], *Wen-bin Gong*[1,2], *Yan-Na Yang*[1,2], *Xin-jie Qi*[1,2] *and Shi Zhang*[1,2*]

[1] *Department of Pulmonary and Critical Care Medicine, Jinan Central Hospital, Cheeloo College of Medicine, Shandong University, Jinan, China,* [2] *Department of Pulmonary and Critical Care Medicine, Jinan Central Hospital Affiliated to Shandong First Medical University, Jinan, China*

Object: The fluid management strategy in ARDS is not very clear. A secondary analysis of RCT data was conducted to identify patients with ARDS benefitting from a conservative strategy of fluid management.

Methods: The data of this study were downloaded from the ARDS network series of randomized controlled trials (Conservative Strategy vs. Liberal Strategy in 2006). Based on the clinical feature of patients, within the first 24 h after admission, clustering was performed using the *k*-means clustering algorithm to identify the phenotypes of ARDS. Survival was analyzed using the Kaplan-Meier survival analysis to assess the effect of the two fluid management strategies on the 90-day cumulative mortality. Categorical/dichotomic variables were analyzed by the chi-square test. Continuous variables were expressed as the mean and standard deviation and evaluated through a one-way ANOVA. A P-value < 0.05 was defined as the statistically significant cut-off value.

Results: A total of 1,000 ARDS patients were enrolled in this unsupervised clustering research study, of which 503 patients were treated with a conservative fluid-management strategy, and 497 patients were treated with a liberal fluid-management strategy. The first 7-day cumulative fluid balance in patients with the conservative strategy and liberal strategy were -136 ± 491 ml and $6{,}992 \pm 502$ ml, respectively ($P < 0.001$). Four phenotypes were found, and the conservative fluid-management strategy significantly improved the 90-day cumulative mortality compared with the liberal fluid-management strategy (HR $= 0.532$, $P = 0.024$) in patients classified as "hyperinflammatory anasarca" phenotype (phenotype II). The characteristics of this phenotype exhibited a higher WBC count ($20487.51 \pm 7223.86/mm^3$) with a higher incidence of anasarca (8.3%) and incidence of shock (26.6%) at baseline. The furthermore analysis found that the conservative fluid management strategy was superior to the liberal fluid management strategy in avoiding superinfection (10.10 vs. 14.40%, $P = 0.037$) and returned to assisted breathing (4.60 vs. 16.20%, $P = 0.030$) in patients classified as

OPEN ACCESS

Edited by:
Qinghe Meng,
Upstate Medical University, United States

Reviewed by:
Abele Donati,
Marche Polytechnic University, Italy
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

***Correspondence:**
Shi Zhang
394873967@qq.com

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *20 June 2021*
Accepted: *30 July 2021*
Published: *25 August 2021*

Citation:
Xing C-y, Gong W-b, Yang Y-N, Qi X-j and Zhang S (2021) ARDS Patients Exhibiting a "Hyperinflammatory Anasarca" Phenotype Could Benefit From a Conservative Fluid Management Strategy. Front. Med. 8:727910. doi: 10.3389/fmed.2021.727910

"hyperinflammatory anasarca" phenotype. In addition, patients with other phenotypes given the different fluid management strategies did not show significant differences in clinical outcomes.

Conclusion: Patients exhibiting a "hyperinflammatory anasarca" phenotype could benefit from a conservative fluid management strategy.

Keywords: ARDS, conservative fluid management, liberal fluid management, phenotype, hyperinflammatory anasarca

BACKGROUND

Acute respiratory distress syndrome (ARDS) refers to acute inflammatory injury of the lung, disruption of the alveolar–capillary barrier and the formation of non-cardiogenic, protein-rich pulmonary oedema (1–4). A conservative fluid management strategy could improve the anasarca and oxygenation index (PaO2/FiO2). In addition, initiating treatment to reduce pulmonary oedema as early as possible could decrease the risk of superinfection (5, 6).

Although lung failure alone can be lethal, death in patients with acute lung injury is usually due to the failure of the non-pulmonary organs. Conservative fluid management strategies may lead to lower intravascular volume and perfusion (7, 8). Wiedemann et al. performed a randomized controlled trial (RCT) to compare conservative and liberal fluid-management strategies in ARDS (9). The results indicated that although the conservative strategy of fluid management improved lung function and shortened the duration of mechanical ventilation, a conservative strategy could not improve the mortality of ARDS. This suggested that not all ARDS patients need dehydration therapy for the improvement of lung function, and the sufficient effective circulating blood volume could also be took into account in parts of ARDS patient. Therefore, the fluid management strategy for ARDS is not very clear.

Artificial intelligence (AI) has found its way into clinical studies in the era of big data. Meanwhile, as increasing number of ARDS clinical trials data is open to public, secondary analysis on these combined datasets provide a powerful way of finding solution to clinical questions with a new perspective (10, 11). When combined with machine learning informatics and clinical trials data, the result will be the development of a precision form of personalized treatment applied to ARDS, which could be a promising way to explore the precise fluid management for specific ARDS population (12).

Based on this clinical problem, the hypothesis for identification of the specific ARDS patients who could benefit from conservative fluid management would be tested through a secondary analysis on RCT data from Wiedemann et al. using a machine learning algorithm (unsupervised clustering).

METHODS

The data of this study were downloaded from the ARDS network series of randomized controlled trials (Conservative Strategy vs. Liberal Strategy in 2006) (9). A total of 1,000 ARDS patients participated in this study.

Screening Features for Unsupervised Clustering

Clinical features of ARDS patients were obtained before the start of treatment with a conservative strategy or liberal strategy within the first 24 h after admission. If missing data for a certain feature or sample is more than 5% then we will leave that feature. The other missing data (<5%) were estimated by multiple imputations through the R package, following the process described by Zhang with minor modifications (13). The mice R package conducted three main steps: (1) imputation, (2) analysis, and (3) pooling for missing data. The imputation step identified the characteristic of missing data; then the analysis step provided the predictive mean matching of missing data through modular approach; finally, the pooling step filled up the missing data based on 1,000 imputations iterations (13–15).

To screen suitable clinical features for clustering analysis, we attempted to train several classifiers from scratch. The clear separations and significant statistical results ($P < 0.05$) were utilized as the criterion for the identification of suitable clinical features for the best classification model.

Statistical Methods

Clustering was performed using the *k*-means clustering algorithm implemented in R (*k*-means package). The best classifications were selected based on clear separations of the consensus heatmaps.

Survival was analyzed using the Kaplan-Meier survival analysis to assess the effect of the two fluid management strategies on the 90-day cumulative mortality. Categorical/dichotomic variables were analyzed by the chi-square test. Continuous variables were expressed as the mean and standard deviation and evaluated through a one-way ANOVA.

A P-value < 0.05 was defined as the statistically significant cut-off value.

Software

All the analyses in this study were conducted using R 4.0.3.

RESULTS

Patients

A total of 1,000 ARDS patients were enrolled in this unsupervised clustering research study, of which 503 patients were treated with a conservative fluid-management strategy, and 497 patients were treated with a liberal strategy fluid-management strategy. The first 7-day cumulative fluid balance in patients with the

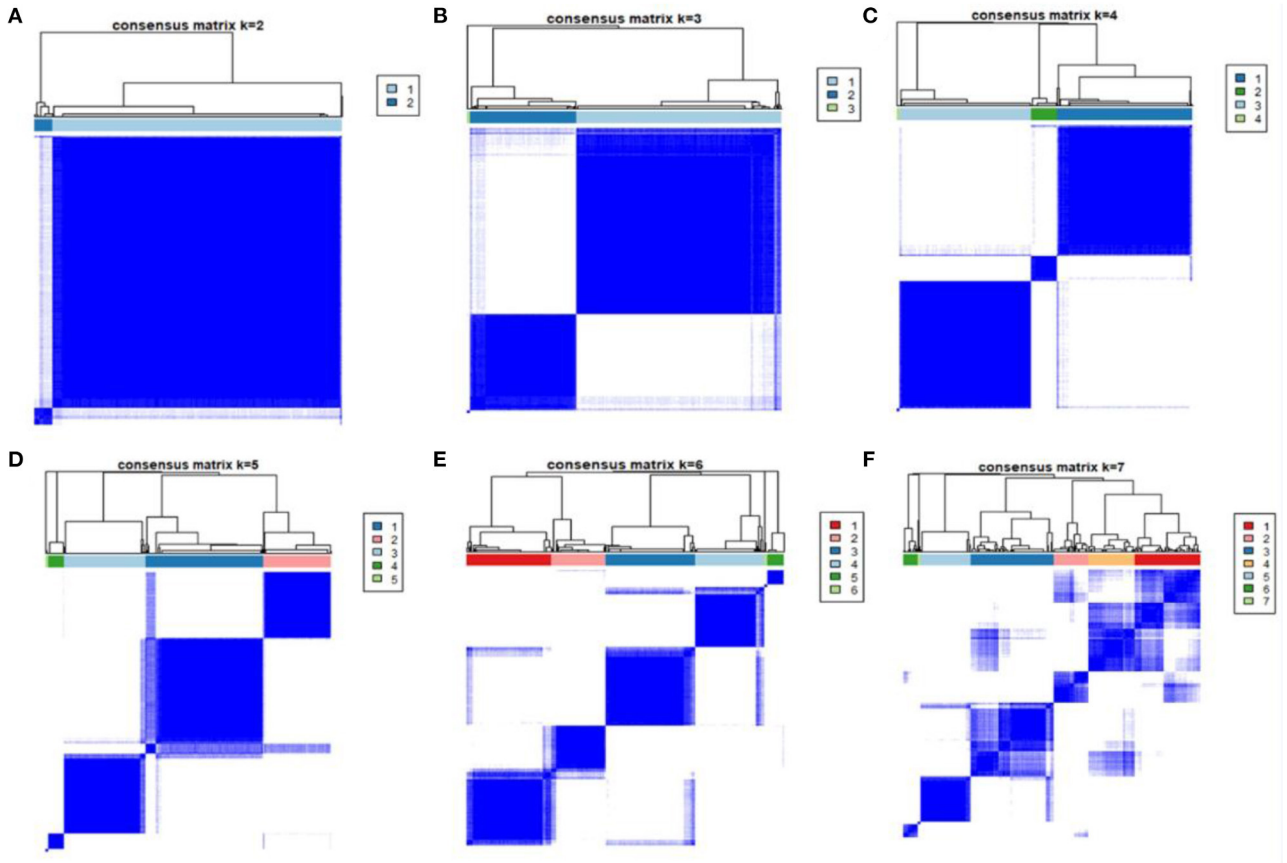


FIGURE 1 | Consensus matrix heatmaps of consensus *k*-means clustering. Consensus matrix heatmaps of different subgroup numbers ($k = 2, 3, 4, 5, 6, 7$). When $k = 3$, the model exhibited the clearest separation of the consensus matrix heatmap.

conservative strategy and liberal strategy were -136 ± 491 ml and $6{,}992 \pm 502$ ml, respectively ($P < 0.001$).

Characteristics for Unsupervised Clustering

After multiclustering, the Acute Physiology and Chronic Health Evaluation III (APACHE III) score, PaO2, central venous pressure (CVP), predicted body weight (PBW), white blood cell count (WBC), platelet count, and the presence or absence of shock and anasarca were finally enrolled in further unsupervised clustering analysis.

Clinical Outcomes of Phenotypes

The patients were classified as 2 phenotypes to 7 phenotypes through unsupervised clustering analysis, shown in (**Figures 1A–F**). As the 4-class model showed the clearest separation of the matrix heatmap (**Figure 1**), 4 phenotypes were utilized in the current study. The numbers of patients in Phenotypes I, II, III and IV were 319, 169, 492 and 11, respectively.

Phenotype II was identified as the specific population that benefited from the conservative fluid-management strategy because the conservative fluid-management strategy significantly improved the 90-day cumulative mortality compared with the liberal fluid-management strategy (HR $= 0.532$, $P = 0.024$), as shown in **Figure 2**. Regarding secondary outcomes, the conservative fluid management strategy markedly decreased the 90-day mortality compared with the liberal fluid management strategy (25.3 vs. 41.1%, $P = 0.030$). In addition, the conservative fluid management strategy was superior to the liberal fluid management strategy in avoiding superinfection (10.10 vs. 14.40%, $P = 0.037$) and returned to assisted breathing (4.60 vs. 16.20%, $P = 0.030$), as shown in **Table 1**.

Patients with other phenotypes given the different fluid management strategies did not show a significant difference in clinical outcomes, as shown in **Figure 2** and **Table 1**.

Features of the Phenotypes

For better insight into the characteristics of the phenotypes, features among different phenotypes were compared and evaluated. Phenotype IV was not selected as the main observational cohort due to the small sample size.

Patients classified as phenotype II exhibited a higher WBC ($20487.51 \pm 7223.86/mm^3$) and had a higher incidence of anasarca (8.3%) and incidence of shock (26.6%) at baseline, as shown in **Tables 2**, **3**. Therefore, phenotype II was defined as the

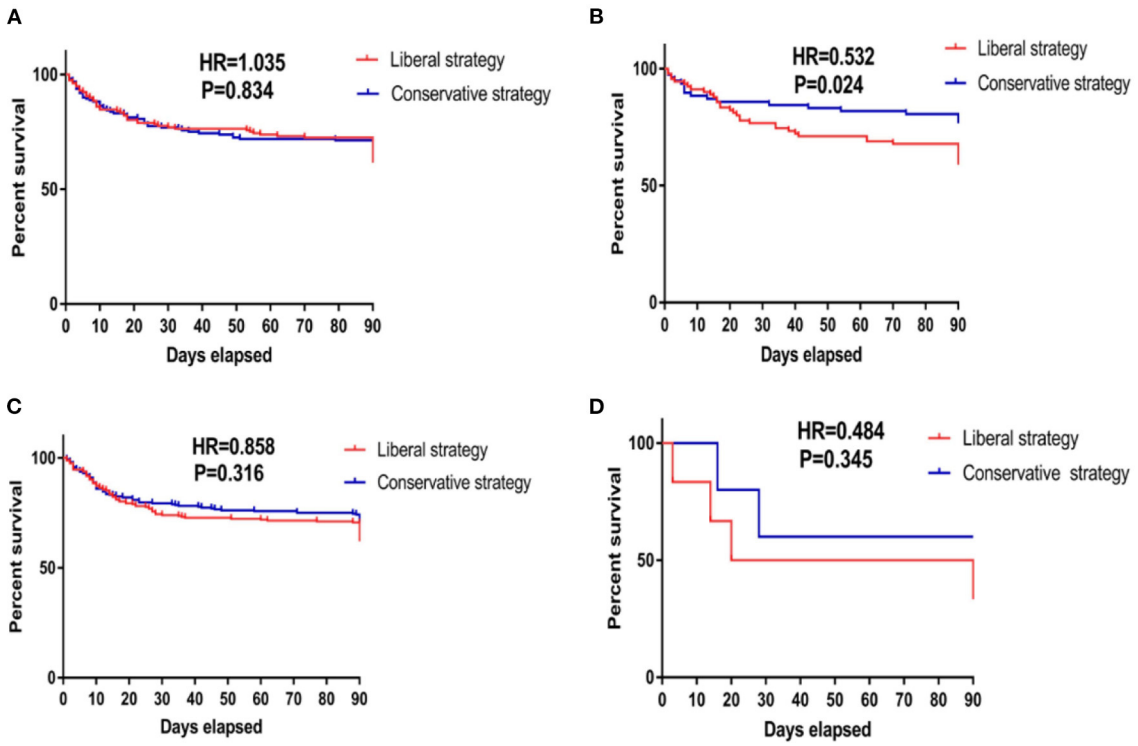


FIGURE 2 | Kaplan–Meier survival curves of 90-day cumulative mortality for patients receiving conservative strategy and liberal strategy among those with the four phenotypes. **(A)** The survival curves of 90-day cumulative mortality of patients classified as phenotype I. The Kaplan–Meier survival analysis indicated that compared with the liberal strategy, conservative fluid management could not improve the 90-day mortality in patients classified as phenotype I (HR = 1.035, $P = 0.843$). **(B)** The survival curves of 90-day cumulative mortality of patients classified as phenotype II. The Kaplan–Meier survival analysis indicated that compared with the liberal strategy, conservative fluid management significantly improved 90-day mortality in patients classified as phenotype II (HR = 0.532, $P = 0.024$). **(C)** The survival curves of 90-day cumulative mortality of patients classified as phenotype III. The Kaplan–Meier survival analysis indicated that compared with the liberal strategy, conservative fluid management could not improve the 90-day mortality in patients classified as phenotype III (HR = 0.858, $P = 0.316$). **(D)** The survival curves of 90-day cumulative mortality of patients classified as phenotype IV. The Kaplan–Meier survival analysis indicated that compared with the liberal strategy, conservative fluid management could not improve the 90-day mortality in patients classified as phenotype IV (HR = 0.484, $P = 0.345$). HR, hazard ratio.

"hyperinflammatory anasarca" phenotype. Other characteristics of phenotypes are illuminated in **Tables 2**, **3**.

DISCUSSION

The fluid management strategy for ARDS is not very clear. The current secondary analysis of RCTs identified 4 ARDS phenotypes, and a conservative fluid management strategy significantly improved the 90-day mortality of patients classified as phenotype II compared with a liberal fluid management strategy. In addition, a conservative fluid-management strategy was superior to a liberal fluid-management strategy in avoiding superinfection and returned to assisted breathing. Phenotype II was defined as a "hyperinflammatory anasarca" phenotype due to the higher WBC count with the higher incidence of anasarca and incidence of shock at baseline.

The current study first found that patients exhibiting a "hyperinflammatory anasarca" phenotype could benefit from a conservative fluid management strategy. This specific population showed a higher WBC ($20487.51 \pm 7223.86/mm^3$) with a higher incidence of anasarca (8.3%) and incidence of shock (26.6%) at baseline. Distributive shock and oedema due to ARDS-induced systemic inflammatory host responses on cardiovascular systems were marked signs in these patients (16–19). Previous studies uncovered that oedema was an independent risk factor for superinfection. and the anasarca could increase the number of days of mechanical ventilation (2, 20, 21). Our analysis further demonstrated that relieving oedema through a conservative fluid management strategy could effectively avoid superinfection and return to assisted breathing in patients with phenotype II, which could be the main reason to explain why the conservative fluid management strategy improved the mortality of these ARDS populations. Meanwhile, in order to maintain mean arterial pressure $\geq$ 65 mmHg and sufficient cardiac output to achieve adequate tissue perfusion for important organs, vasopressors are critical and should be used early for patients classified as phenotype II. This strategy is also suggested by Surviving Sepsis Campaign guidelines for septic shock (22, 23).

Individual and detailed situations should be considered to select a suitable fluid management strategy in patients classified as having other phenotypes. As there were no significant differences

TABLE 1 | Secondary outcomes in phenotype 2.

Outcomes	Conservative strategy ($n = 90$)	Liberal strategy ($n = 79$)	*P*
28 day mortality (%)	16.50%	23.30%	0.268
60 day mortality (%)	20.30%	32.50%	0.197
90 day mortality (%)	25.30%	41.10%	**0.030**
Unassisted breathing (%)	17.70%	24.40%	0.290
Super infection (%)	10.10%	14.40%	**0.037**
Returned to assisted breathing (%)	4.60%	16.20%	**0.030**
Hospital free days to 90 (day)	51.75 ± 43.50	48.12 ± 34.97	0.503
ICU free days to day 90 (day)	62.69 ± 31.91	57.73 ± 34.22	0.097
Ventilator free days to day 90	61.40 ± 35.24	63.27 ± 32.75	0.310

TABLE 2 | Dichotomous characteristics in different phenotypes.

Id	Phenotype 1 ($n = 319$)	Phenotype 2 ($n = 169$)	Phenotype 3 ($n = 492$)	Phenotype 4 ($n = 11$)	*P*
Male sex	52.3%	56.8%	53.9%	63.6%	0.674
Shock	6.6%	8.3%	2.8%	9.1%	**0.011**
Surgery	4.1%	3.0%	6.1%	0.0%	0.294
Ethanol	9.1%	11.2%	13.4%	18.2%	0.302
ARDS risk factor					
Pneumonia	45.5%	48.5%	48.2%	63.4%	0.187
Sepsis	22.3%	18.9%	25.8%	9.1%	0.385
Trauma	8.5%	7.7%	6.7%	9.1%	0.874
Multiple transfusion	0.9%	0	1.0%	9.1%	0.223
Aspiration	15.7%	15.4%	14.6%	9.1%	0.483
Others	6.6%	9.5%	5.5%	0	0.304
Anamnesis					
AIDS	6.0%	4.7%	9.1%	9.1%	0.200
Leukemia	1.9%	0.6%	3.5%	9.1%	0.085
Lymphoma	0.6%	0.6%	2.0%	0.0%	0.282
Solid tumor	2.2%	0.0%	1.8%	0.0%	0.289
Immune suppression	4.7%	9.5%	10.6%	9.1%	**0.037**
Anasarca	16.0%	26.6%	17.7%	18.2%	**0.027**
Heart failure	4.7%	4.7%	2.8%	9.1%	0.343
Hypertension	32.9%	34.3%	28.5%	18.2%	0.224
Myocardial infarction	6.9%	5.9%	4.9%	0.0%	0.510
Dementia	2.2%	3.6%	2.2%	0.0%	0.713
Stroke	4.1%	3.6%	5.1%	0%	0.732
Hepatic failure	0.9%	0.5%	1.0%	0.0%	0.953
Cirrhosis	2.2%	2.4%	4.5%	0.0%	0.270
Peptic ulcer	5.6%	4.7%	3.0%	9.1%	0.238
Diabetes	18.8%	18.9%	17.5%	0.0%	0.409

in clinical outcomes between conservative and liberal fluid management strategies, the detailed therapies should depend on patients' individual morbid conditions. If shock-induced tissue hypoperfusion is a crucial clinical problem in certain patients, a conservative fluid management strategy should be cautiously used in these patients. However, if ARDS-induced shock is reversed, a conservative fluid-management strategy could be considered for the improvement of respiratory failure (24–26).

There are some limitations in this study: prospective validation is required before definitive conclusions regarding therapy can be drawn. Meanwhile, the study is also limited by the fact that the beneficial effect was not externally validated. In addition, specific populations who could benefit from liberal fluid management strategies or other therapeutic methods could not be identified in the current study.

TABLE 3 | Continuous variables in different phenotypes.

Id	Phenotype 1	Phenotype 2	Phenotype 3	Phenotype 4	*P*
Age (year)	48.57 ± 16.15	53.86 ± 15.96	50.50 ± 15.88	47.82 ± 16.88	0.391
Height (cm)	169.14 ± 11.53	169.46 ± 10.57	170.19 ± 10.23	175.56 ± 5.54	0.059
Weight	83.32 ± 24.57	81.96 ± 23.07	81.91 ± 22.53	89.92 ± 18.89	0.588
Temperature (°C)	37.4 ± 0.9	37.4 ± 0.9	37.6 ± 1.1	37.6 ± 1.4	0.136
Systolic BP (mmHg)	113.19 ± 20.88	113.47 ± 19.86	114.09 ± 23.10	119.82 ± 19.21	0.751
Diastolic BP (mmHg)	59.58 ± 12.13	59.31 ± 12.78	59.30 ± 13.15	62.64 ± 12.43	0.847
Mean arterial pressure (mmHg)	77.17 ± 14.04	77.22 ± 13.79	77.04 ± 14.49	81.92 ± 13.50	0.737
Heart rate (bpm)	102.05 ± 20.42	100.02 ± 20.10	103.17 ± 21.83	98.09 ± 19.71	0.349
Respiratory rate (breaths/min)	34.31 ± 9.43	34.66 ± 9.03	35.14 ± 10.67	35.55 ± 17.15	0.712
CVP (mm H_2O)	11.79 ± 4.61	11.89 ± 4.56	12.19 ± 4.90	11.55 ± 4.78	0.661
Urine output/24 h (ml)	1978.81 ± 1348.99	2150.12 ± 2063.27	2155.51 ± 1633.35	1908.27 ± 1212.69	0.448
Glasgow coma	10.59 ± 4.58	10.71 ± 4.50	10.91 ± 4.40	9.82 ± 4.24	0.689
PaO2 (mmHg)	92.40 ± 44.30	94.54 ± 43.55	96.04 ± 44.76	119.91 ± 74.04	0.191
PaO2/FiO2	125.93 ± 61.51	133.80 ± 66.92	133.84 ± 61.26	164.30 ± 77.28	0.091
Bicarbonate (mEq/L)	21.68 ± 5.43	21.67 ± 5.88	20.89 ± 5.61	21.00 ± 5.27	0.178
HCT (%)	32.74 ± 7.27	32.6 ± 6.69	32.29 ± 6.76	30.82 ± 4.51	0.677
Glucose (mg/dL)	142.92 ± 88.07	149.83 ± 79.36	135.78 ± 58.75	113.36 ± 20.87	0.080
Potassium (mEq/L)	3.98 ± 0.65	4.06 ± 0.66	3.98 ± 0.64	4.46 ± 0.60	**0.045**
Sodium (mEq/L)	138.74 ± 5.17	139.31 ± 6.86	138.88 ± 5.16	138.27 ± 3.74	0.717
WBC (/mm^3)	17313.76 ± 10409.86	20487.51 ± 7223.86	8402.11 ± 9878.86	66836.36 ± 31126.13	**<0.001**
HGB (g/dL)	10.66 ± 2.04	10.63 ± 1.93	10.18 ± 1.81	10.05 ± 0.90	**0.001**
Platelets (1,000/mm^3)	226.44 ± 123.08	244.66 ± 147.44	158.12 ± 106.95	135.36 ± 54.93	**<0.001**
Albumin (g/dL)	2.24 ± 0.66	2.16 ± 0.63	2.19 ± 0.62	2.50 ± 0.73	0.250
Bilirubin (mg/dL)	1.56 ± 2.86	1.48 ± 1.78	1.78 ± 4.21	3.76 ± 7.39	0.162
BUN (mg/dL)	22.74 ± 16.81	20.09 ± 23.37	25.25 ± 19.08	21.18 ± 10.85	0.025
Chloride (mEq/L)	107.43 ± 6.81	106.70 ± 8.32	108.15 ± 6.37	108.18 ± 5.74	0.100
Creatinine (mg/dL)	1.16 ± 0.77	1.28 ± 0.81	1.33 ± 0.92	1.14 ± 0.40	**0.039**
Total protein	5.11 ± 1.04	5.05 ± 0.99	4.95 ± 1.04	5.28 ± 0.94	0.136
APACHE III	95.72 ± 32.75	92.07 ± 27.18	93.99 ± 30.98	87.27 ± 25.56	0.551

WBC, white blood cell count; HCT, red blood cell specific volume; HGB, hemoglobin concentration; BUN, blood urea nitrogen. Bold values mean P value less than 0.05.

CONCLUSIONS

Patients exhibiting a "hyperinflammatory anasarca" phenotype could benefit from conservative fluid management strategies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: ARDS network public database (http://www.ardsnet.org/).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

SZ took responsibility for the integrity and the accuracy of the data analysis. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

Thank you for all the researchers of the ARDS network project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.727910/full#supplementary-material

REFERENCES

1. Thompson BT, Chambers RC, Liu KD. Acute respiratory distress syndrome. *N Engl J Med.* (2017) 377:1904–5. doi: 10.1056/NEJMra1608077
2. Fan E, Brodie D, Slutsky AS. Acute respiratory distress syndrome: advances in diagnosis and treatment. *JAMA.* (2018) 319:698–710. doi: 10.1001/jama.2017.21907
3. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA.* (2016) 315:788–800. doi: 10.1001/jama.2016.0291
4. Bos LD, Martin-Loeches I, Schultz MJ. ARDS: challenges in patient care and frontiers in research. *Eur Respir Rev.* (2018) 27:170107. doi: 10.1183/16000617.0107-2017
5. Vignon P, Evrard B, Asfar P, Busana M, Calfee CS, Coppola S, et al. Fluid administration and monitoring in ARDS: which management? *Intensive Care Med.* (2020) 46:2252–64. doi: 10.1007/s00134-020-06310-0
6. Casey JD, Semler MW, Rice TW. Fluid management in acute respiratory distress syndrome. *Semin Respir Crit Care Med.* (2019) 40:57–65. doi: 10.1055/s-0039-1685206
7. Silversides JA, Major E, Ferguson AJ, Mann EE, McAuley DF, Marshall JC, et al. Conservative fluid management or deresuscitation for patients with sepsis or acute respiratory distress syndrome following the resuscitation phase of critical illness: a systematic review and meta-analysis. *Intens Care Med.* (2017) 43:155–70. doi: 10.1007/s00134-016-4573-3
8. Monnet X, Teboul JL. My patient has received fluid. How to assess its efficacy and side effects? *Ann Intens Care.* (2018) 8:54. doi: 10.1186/s13613-018-0400-z
9. Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med.* (2006) 354:2564–75. doi: 10.1016/j.jvs.2006.08.053
10. Zhang Z, Navarese EP, Zheng B, Meng Q, Liu N, Ge H, et al. Analytics with artificial intelligence to advance the treatment of acute respiratory distress syndrome. *J Evid Based Med.* (2020) 13:301–12. doi: 10.1111/jebm.12418
11. Zhang Z, Zheng B, Liu N, Ge H, Hong Y. Mechanical power normalized to predicted body weight as a predictor of mortality in patients with acute respiratory distress syndrome. *Intens Care Med.* (2019) 45:856–64. doi: 10.1007/s00134-019-05627-9
12. Bland JS. The natural roots of functional medicine. *Integr Med.* (2018) 17:12–7.
13. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. *Ann Transl Med.* (2016) 4:30. doi: 10.3978/j.issn.2305-5839.2015.12.63
14. Zhang S, Lu Z, Wu Z, Xie J, Yang Y, Qiu H. Determination of a "specific population who could benefit from rosuvastatin": a secondary analysis of a randomized controlled trial to uncover the novel value of rosuvastatin for the precise treatment of ARDS. *Front Med.* (2020) 7:598621. doi: 10.3389/fmed.2020.598621
15. Qin X, Li J, Hu W, Yang J. Machine learning K-means clustering algorithm for interpolative separable density fitting to accelerate hybrid functional calculations with numerical atomic orbitals. *J Phys Chem A.* (2020) 124:10066–174. doi: 10.1021/acs.jpca.0c06019
16. Zhang S, Chu C, Wu Z, Liu F, Xie J, Yang Y, et al. IFIH1 contributes to M1 macrophage polarization in ARDS. *Front Immunol.* (2021) 11:580838. doi: 10.3389/fimmu.2020.580838
17. Meduri GU, Annane D, Chrousos GP, Marik PE, Sinclair SE. Activation and regulation of systemic inflammation in ARDS: rationale for prolonged glucocorticoid therapy. *Chest.* (2009) 136:1631–43. doi: 10.1378/chest.08-2408
18. Zhang S, Wu Z, Xie J, Yang Y, Wang L, Qiu H. DNA methylation exploration for ARDS: a multi-omics and multi-microarray interrelated analysis. *J Transl Med.* (2019) 17:345. doi: 10.1186/s12967-019-2090-1
19. Huppert LA, Matthay MA, Ware LB. Pathogenesis of acute respiratory distress syndrome. *Semin Respir Crit Care Med.* (2019) 40:31–9. doi: 10.1055/s-0039-1683996
20. Chiumello D, Brioni M. Severe hypoxemia: which strategy to choose. *Crit Care.* (2016) 20:132. doi: 10.1186/s13054-016-1304-7
21. Famous KR, Delucchi K, Ware LB, Kangelaris KN, Liu KD, Thompson BT, et al. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med.* (2017) 195:331–8. doi: 10.1164/rccm.201603-0645OC
22. Coopersmith CM, De Backer D, Deutschman CS, Ferrer R, Lat I, Machado FR, et al. Surviving sepsis campaign: research priorities for sepsis and septic shock. *Intens Care Med.* (2018) 44:1400–26. doi: 10.1007/s00134-018-5175-z
23. Zhang S, Liu F, Wu Z, Xie J, Yang Y, Qiu H. Contribution of m6A subtype classification on heterogeneity of sepsis. *Ann Transl Med.* (2020) 8:306. doi: 10.21037/atm.2020.03.07
24. Evans RG, Naidu B. Does a conservative fluid management strategy in the perioperative management of lung resection patients reduce the risk of acute lung injury? *Interact Cardiovasc Thorac Surg.* (2012) 15:498–504. doi: 10.1093/icvts/ivs175
25. Machare-Delgado E, Decaro M, Marik PE. Inferior vena cava variation compared to pulse contour analysis as predictors of fluid responsiveness: a prospective cohort study. *J Intens Care Med.* (2011) 26:116–24. doi: 10.1177/0885066610384192
26. Peck TJ, Hibbert KA. Recent advances in the understanding and management of ARDS. *F1000Res.* (2019) 8:F1000. doi: 10.12688/f1000research.20411.1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 21 September 2021
doi: 10.3389/fmed.2021.740675

The Effect of Loop Diuretics on 28-Day Mortality in Patients With Acute Respiratory Distress Syndrome

Rui Zhang[1†], *Hui Chen*[1,2†], *Zhiwei Gao*[1,3], *Meihao Liang*[1], *Haibo Qiu*[1], *Yi Yang*[1] *and Ling Liu*[1*]

[1] *Jiangsu Provincial Key Laboratory of Critical Care Medicine, Department of Critical Care Medicine, School of Medicine, Zhongda Hospital, Southeast University, Nanjing, China,* [2] *Department of Critical Care Medicine, The First Affiliated Hospital of Soochow University, Suzhou, China,* [3] *Department of Critical Care Medicine, Huai'an First People's Hospital, Nanjing Medical University, Huai'an, China*

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Hiroshi Morimatsu,
Okayama University, Japan
Ling Sang,
First Affiliated Hospital of Guangzhou Medical University, China
Bo Hu,
Zhongnan Hospital of Wuhan University, China

***Correspondence:**
Ling Liu
liulingdoctor@126.com

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *13 July 2021*
Accepted: *19 August 2021*
Published: *21 September 2021*

Citation:
Zhang R, Chen H, Gao Z, Liang M, Qiu H, Yang Y and Liu L (2021) The Effect of Loop Diuretics on 28-Day Mortality in Patients With Acute Respiratory Distress Syndrome. Front. Med. 8:740675. doi: 10.3389/fmed.2021.740675

Background: Diuretics have been widely used in critically ill patients while it remains uncertain whether they can reduce mortality in patients with acute respiratory distress syndrome (ARDS). This study aimed to investigate the associations between diuretics and 28-day mortality in patients with ARDS.

Methods: This is a secondary analysis of the ARDS Network Fluid and Catheter Treatment Trial (FACTT) of National Heart, Lung, and Blood Institute. Those patients who did not receive renal replacement therapy within the first 48 h after enrollment in the FACTT were included in the analysis. A marginal structural Cox model (MSCM) was used to investigate the associations between diuretics and 28-day mortality after correction of both the baseline and time-varying variables. The latent class analysis (LCA) and subgroup analysis were performed to identify the kind of patients that could be benefited from diuretics.

Results: A total of 932 patients were enrolled, i.e., 558 patients in the diuretics group and 374 patients in the no diuretics group within the first 48 h. The 28-day mortality was lower in the diuretics group (15.1 vs. 28.1%, $p < 0.001$). In MSCM, diuretics use was related to the improved 28-day mortality (*HR* 0.78; 95% *CI* 0.62–0.99; $p = 0.04$). LCA identified three subtypes, and diuretics were associated with reduced mortality in subtype 3, which was characterized by worse renal function and higher central venous pressure (CVP). A subgroup analysis indicated survival advantage among the female patients, sepsis induced ARDS, and those with the ratio of partial pressure of oxygen to the fractional concentration of inspired oxygen (PaO_2/FiO_2) $\leq$ 150 mmHg, and mean arterial pressure (MAP) $\geq$ 65 mmHg.

Conclusion: Loop diuretics were associated with the reduced 28-day mortality in the patients with ARDS, after controlling for time-varying confounders. Randomized trials are required to verify the association.

Keywords: acute respiratory distress syndrome, diuretics, mortality, marginal structural cox model, subtype

INTRODUCTION

Acute respiratory distress syndrome (ARDS) that results from various insults is associated with a high hospital mortality rate of 40% (1). The hallmark alteration in ARDS is increased endothelial and epithelial permeability, leading to the increased extravascular lung water (EVLW) (2), which is associated with lung injury and mortality (3). Diuretics are frequently administered to critically ill patients to alleviate pulmonary edema and may reduce lung injury (4).

Several studies have involved diuretics as part of therapeutic intervention for ARDS, but whether they could reduce mortality has not been conclusively determined. Diuretics have been associated with reduced positive fluid balance, improved lung function, and shorter mechanical ventilation duration, but no significant improvement in the mortality rate has been demonstrated (5, 6). One retrospective study suggested that the use of diuretics for 48–72 h after meeting the ARDS criteria may reduce mortality (7). In that study, the influence of diuretics use beyond the specified 24 h was not analyzed, and nor were therapeutic changes related to diuretics, rendering the result less explicable.

There are several theoretical reasons for the controversial results reported to date. ARDS is of extreme heterogeneity, and patients with diverse phenotypes respond differently to the selected treatment (8–10). In addition, the use of diuretics is a time-dependent variable that is affected by factors, such as oxygenation and mean arterial pressure, but these confounders were seldom corrected in studies, leading to bias. In a previous study, after adjusting for the baseline variables only, diuretics were associated with lower 28-day mortality in critically ill patients. When time-varying confounders were corrected *via* marginal structural Cox modeling (MSCM), however, there was no significant association (11). The study highlighted the necessity to consider time-dependent variables when investigating the effect of diuretics on patient outcomes.

Diuretics are widely used in critically ill patients, despite controversy with respect to whether they reduce mortality. The present study aimed to investigate the effects of loop diuretics on 28-day mortality in the patients with ARDS, and used a marginal structural model to adjust time-varying covariates. We hypothesized that diuretics would improve 28-day mortality in patients with ARDS. A latent class analysis (LCA) was used to derive phenotypes, and subgroup analysis was conducted to determine the phenotypes that may benefit from diuretics.

METHODS

Study Design and Population

The study was a secondary analysis of the ARDS Network Fluid and Catheter Treatment Trial (FACTT) of the National Heart, Lung, and Blood Institute. The details of the trial have been published previously (5, 12). In the original study, the patients with ARDS who received mechanical ventilation were included. The patients with ARDS for more than 48 h, chronic diseases that impair survival and weaning were excluded from the study. We further excluded patients receiving renal replacement therapy routinely or within the first 48 h after enrollment, to whom diuretics were not likely to be prescribed.

Fluid management strategies were conducted for 7 days from randomization, or until weaning, whichever occurred first. Furosemide or other diuretics were administered to the patients with elevated central venous pressure (CVP) or pulmonary arterial wedge pressure when hemodynamics was stable. Patients were divided into two groups according to whether they received diuretics within the first 48 h. The primary outcome was the 28-day mortality. All data were obtained and approved by Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC, https://biolincc.nhlbi.nih.gov). The present study was approved by the Research Ethics Commission of Zhongda Hospital, School of Medicine, Southeast University (Nanjing, China). The Strengthening the Reporting of Observational Studies in the Epidemiology (STROBE) recommendations were followed in this study.

Data Collection

The data extracted included demographic data, laboratory tests, Acute Physiology and Chronic Health Evaluation III score (APACHE III), and prescriptions of vasopressor and diuretics. Sequential Organ Failure Assessment (SOFA) score, Charlson Comorbidity Index (13), and Murray lung injury score (14) were calculated. The number of missing or censoring values is presented in **Supplementary Table 1**. Variables with a missing ratio of more than 25% were not included in the final analysis. Outliers were censored and missing values were replaced by multiple imputations.

Statistical Analysis

The continuous variables were presented as mean (SD) or median [interquartile ranges (IQR) and were compared with Student's *t*-test or the Mann–Whitney test. Categorical variables were compared via the chi-square test or Fisher's exact test. Standardized mean differences (SMDs) and *p*-values were calculated to evaluate the differences between the two groups.

Latent class analysis was employed to derive phenotypes. Variables were selected based on the previous research and potential association with outcomes (9, 15), such as demographic parameters (gender, age, and BMI), comorbidities (diabetes, hypertension, and heart failure), disease severity (APACHE III), vital signs (heart rate, temperature, and respiratory rate), hemodynamic parameters (MAP, CVP, vasopressor use, and fluid balance), respiratory variables (tidal volume and plateau pressure), hematology (platelet and hemoglobin), and the renal function indicator creatinine. Mplus (version 8.3) software was used to fit models with latent classes. The optimal number of classes was determined by a combination of Bayesian information criterion (BIC), entropy, and the Vuong-Lo-Mendell-Rubin (VLMR) test (16).

Abbreviations: HR, hazard ratio; CI, confidence of interval; APACHE III, acute physiology and chronic health evaluation III; SOFA, sequential organ failure assessment; CVP, central venous pressure; MAP, mean arterial pressure; PEEP, positive end expiration pressure; Pplat, plateau pressure; LOS, length of stay; ICU, intensive care unit; VFDs, ventilation free days; MSCM, marginal structural cox model; LCA, latent class analysis.

TABLE 1 | Demographic and clinical characteristics of patients in two groups.

	No diuretics	Diuretics	*P*	SMD
N	374	558		
Age, years	50.20 (17.20)	49.58 (15.09)	0.561	0.038
Female (*n*, %)	172 (46.0)	261 (46.8)	0.866	0.016
Body mass index	28.93 (7.51)	28.77 (6.54)	0.801	0.023
APACHE III	97.81 (30.41)	88.20 (28.90)	<0.001	0.324
SOFA	8.57 (2.92)	7.35 (2.38)	<0.001	0.455
Primary lung injury (*n*, %)			<0.001	0.354
Sepsis	109 (29.1)	97 (17.4)		
Trauma	28 (7.5)	45 (8.1)		
Aspiration	52 (13.9)	94 (16.8)		
Pneumonia	171 (45.7)	267 (47.8)		
Other	14 (3.7)	55 (9.9)		
Comorbidity (*n*, %)				
Immune suppression	27 (7.2)	44 (7.9)	0.803	0.025
Diabetes	62 (16.6)	93 (16.7)	1.000	0.002
Hypertension	82 (21.9)	139 (24.9)	0.331	0.071
Prior myocardial infarction	11 (2.9)	23 (4.1)	0.445	0.064
Congestive heart failure	10 (2.7)	17 (3.0)	0.894	0.022
Chronic pulmonary disease	32 (8.6)	29 (5.2)	0.058	0.133
Charlson Comorbidity Index	0.00 (0.00, 2.00)	0.00 (0.00, 2.00)	0.077	0.167
Heart rate, bpm	97.65 (19.95)	97.11 (19.41)	0.686	0.027
Respiratory rate, bpm	28.52 (7.17)	27.39 (7.51)	0.023	0.154
CVP, mmHg	11.53 (4.81)	11.71 (5.05)	0.601	0.036
MAP, mmHg	74.89 (13.39)	80.32 (13.54)	< 0.001	0.404
Vasopressors (*n*, %)	172 (46.0)	100 (17.9)	< 0.001	0.631
Fluid balance, ml	2,810.77 (3,261.29)	819.08 (2,802.81)	<0.001	0.655
Hemoglobin, g/dl	9.77 (1.65)	10.00 (1.62)	0.037	0.140
Platelets, $\times 10^{12}$/L	179.63 (121.63)	200.17 (120.25)	0.013	0.170
Creatinine, mg/dl	1.35 (0.99)	1.25 (1.01)	0.151	0.098
PaO_2/FiO_2, mmHg	145.56 (72.57)	149.14 (60.19)	0.441	0.054
Clinical outcomes				
28-day mortality (*n*, %)	105 (28.1)	84 (15.1)	< 0.001	0.321
VFDs by day 28, day	19.00 (14.00, 23.00)	21.00 (17.25, 24.00)	< 0.001	0.315
RRT by day 90 (*n*, %)	18 (7.8)	27 (11.6)	0.215	0.130
RRT days by day 90, day	15.00 (9.00, 32.00)	13.50 (9.00, 28.50)	0.692	0.086

Data are presented as mean (SD), median (interquartile range [IQR]), or number (proportion). APACHE III, acute physiology and chronic health evaluation III; SOFA, sequential organ failure assessment; CVP, central venous pressure; MAP, mean arterial pressure; PaO_2/FiO_2, ratio of partial pressure of oxygen to the fractional concentration of inspired oxygen; VFDs, ventilation free days; RRT, renal replacement therapy; SMD, standardized mean difference.

Daily fluid balance, MAP, need for vasopressors, and the ratio of partial pressure of oxygen to the fractional concentration of inspired oxygen (PaO_2/FiO_2), which would influence the decision of diuretics treatment and potentially correlate with the outcomes, were defined as time-dependent variables. The marginal structural model uses inverse probability of treatment-weighting (IPTW) estimator to create a pseudo-population, enabling the correction of time-fixed baselines and time-varying confounders (17, 18). MSCM was used to evaluate the effect of diuretics on 28-day mortality. Several specified subgroup analyses were performed. In this study, we used RStudio (version 1.3.1073, RStudio Inc., MA, USA) software to perform the statistical analyses. The variable $p < 0.05$ was deemed to indicate the statistical significance.

RESULTS

Demographic and Clinical Characteristics

A total of 932 patients were included in the analysis, of which 558 (59.9%) received diuretics within the first 48 h since enrollment. The demographic and clinical characteristics of patients in the two groups are shown in **Table 1**. In general, patients in the diuretics group had less severe disease, higher mean arterial pressure, and a lower proportion of vasoactive agents use than those in the non-diuretics group. All-cause 28-day mortality was

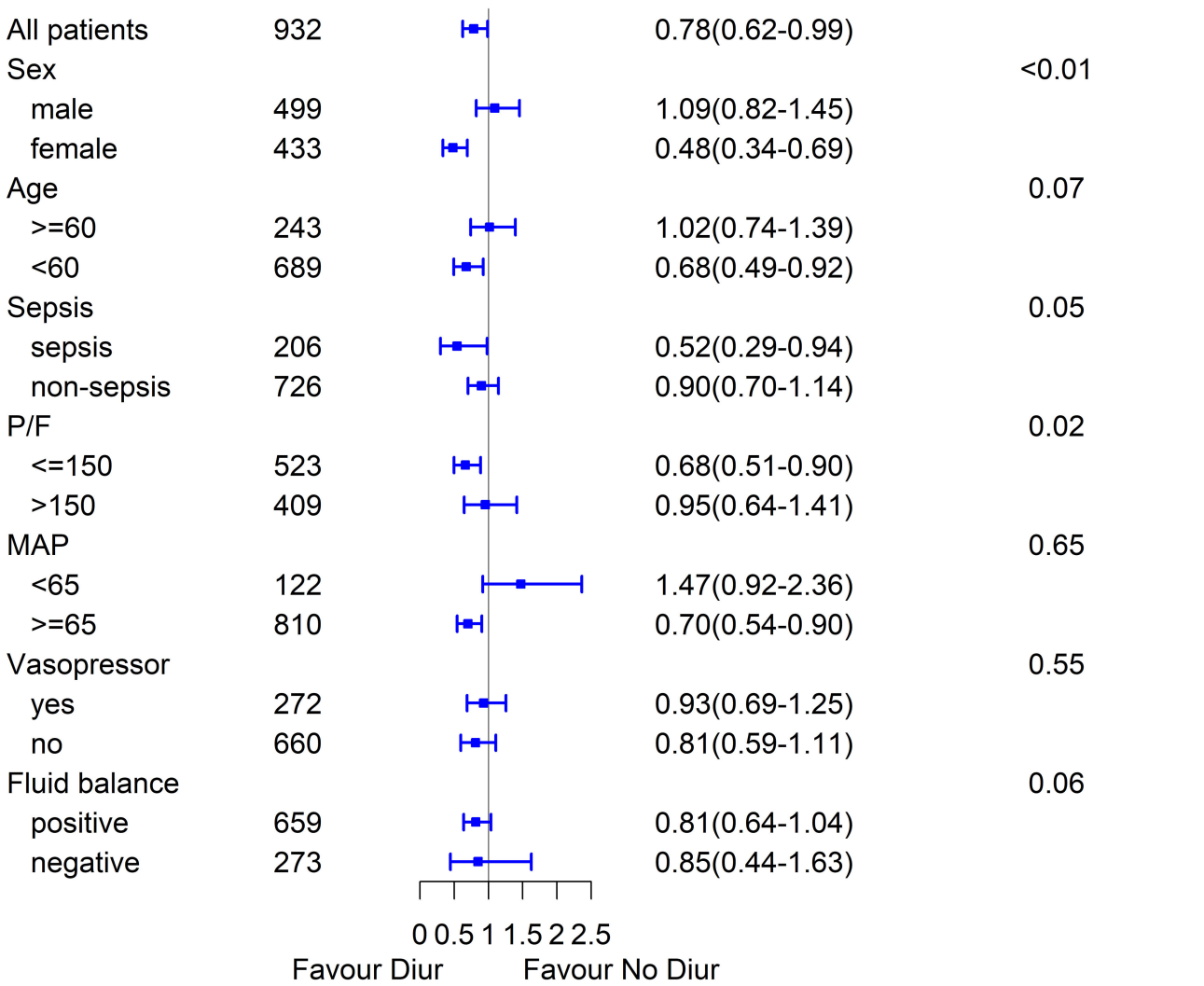


FIGURE 1 | Marginal structural Cox model hazard ratio (HR) values for 28-day mortality in diuretics and no diuretics group according to the subgroups.

significantly lower in the diuretics group [84 (15.1%) vs. 105 (28.1%), $p < 0.001$], and a survival advantage was still evident at day 90. The detailed comparisons between the two groups are presented in **Supplementary Table 2**.

Association Between Diuretics Use and 28-Day Mortality

Time-fixed variables, such as age, APACHE III, Charlson Comorbidity Index, and the time-varying confounders (as mentioned above) were adjusted via marginal structural model. The weight distribution of IPTW applied to adjust for the confounding factors is shown in **Supplementary Figure 1**. Ultimately, the MSCM analysis revealed that compared with no diuretics therapy, loop diuretics use was associated with improved 28-day mortality in the patients with ARDS (HR 0.78; 95% CI 0.62–0.99; $p = 0.04$) in the overall population (**Figure 1**).

The fit statistics of the LCA models generated are shown in **Table 2**. Three main phenotypes were identified, designated as subtypes 1–3. Subtype 1 included 89 (9.5%) patients that mainly suffered from pneumonia, and exhibited relatively normal renal function and the lowest CVP. Subtype 2 included 635 (68.1%) patients who were characterized by near normal serum creatinine and relatively lower CVP. Subtype 3 included 208 (22.3%) patients characterized by worse renal function, higher CVP, and higher proportions of complications, such as diabetes, hypertension, and chronic heart failure. Comparisons among the subtypes 1–3 are shown in **Table 3**, **Supplementary Table 3**, and **Supplementary Figure 2**. MSCM indicated that subtype three patients could be benefited

TABLE 2 | Fit statistics for latent class analysis models.

	BIC	Entropy	*P*-values	Number of patients in each phenotype			
				1	2	3	4
1	51,833			932			
2	51,009	0.746	0.10	604	328		
3	50,069	0.988	0.0001	89	635	208	
4	49,635	0.902	0.03	89	213	424	206

BIC, Bayesian information criterion.

from diuretics (HR 0.64; 95% CI 0.44–0.92; $p = 0.02$), whereas there were no significant associations between diuretics and 28-day mortality in subtype 1 or subtype 2 patients (**Figure 2**).

In the subgroup analysis, diuretics use was correlated with the reduced 28-day mortality in patients with initial MAP equal or more than 65 mmHg, and patients with PaO_2/FiO_2 equal or less than 150 mmHg, and no interaction was detected. Besides, the association seemed to be stronger in female patients with ARDS and sepsis-induced ARDS, and the interaction was significant. Other results of the subgroup analyses are shown in **Figure 1** and **Supplementary Table 4**.

DISCUSSION

In the present study, early loop diuretics were associated with reduced 28-day mortality in patients with ARDS after adjustment for both time-fixed and time-varying confounders. LCA identified three phenotypes and patients in subtype 3 who were characterized by worse renal function and higher CVP, may benefit from diuretics. Additional subgroup analyses of the patients with ARDS indicated that associations between diuretics and reduced 28-day mortality were more marked in female patients, sepsis-induced ARDS, and patients with lower PaO_2/FiO_2 ($\leq$150 mmHg), higher MAP ($\geq$65 mmHg).

Fluid therapy is the fundamental treatment for ARDS, but volume overload is quite common and is associated with an increased risk of death (19). Diuretics are frequently prescribed to the critically ill patients to facilitate liquid removal and have become a pharmacologic adjuvant therapy in patients with ARDS (20). Studies indicate that compared with a liberal fluid strategy or standard care, conservative fluid management achieved by restricting fluid intake and the use of diuretics or hemofiltration is associated with improved oxygenation, increased ventilation-free days, and lower mortality (21–24). It has been proposed that correction of fluid retention may rely on diuretics or renal replacement therapy after the hemodynamic status is stabilized (25). Notably, early diuretics use was independently associated with lower mortality, which had been reported in a less rigorous study that used logistics regression based on the time-fixed baseline variables (7). The effects of diuretics on 28-mortality identified via the use of MSCM to adjust for time-dependent confounders further support their use in patients with ARDS.

There are evident distinctions in the etiology, physiology, and biology of patients with ARDS, leading to different responses to the same therapy (10). Three subtypes were identified by LCA in the present study, and MSCM indicated that diuretics correlated with reduced 28-day mortality in subtype 3, in which patients were characterized by elevated serum creatinine, higher CVP, and more complications, such as diabetes mellitus, hypertension, and heart disease. In a previous study, in patients with ARDS especially with concomitant acute kidney injury, positive fluid balance was associated with higher mortality (26). When used appropriately, however, frusemide may prevent and even resolve acute kidney injury as well as improving survival (27, 28). In another study, diuretics were significantly associated with lower mortality in the positive fluid balance subgroup but there was no significant association in the negative fluid balance subgroup (29). Moreover, diuretics have been recommended in patients with hypertension and heart failure to promote water and sodium excretion and reduce volume load (30, 31). The effects of diuretics on mortality might be attributed to the improvement of renal function and reduction of fluid retention.

Fluid resuscitation is highly recommended in sepsis management (32) but persistent positive fluid balance is an independent risk factor for death (33). Actually, in patients with ARDS complicated by septic shock, achieving both early goal-directed cardiovascular resuscitation and late conservative fluid therapy was reportedly associated with reduced mortality (34). A conservative fluid strategy has been recommended for sepsis-induced ARDS in which there is no evidence of tissue hypoperfusion (32). The associations between diuretics and reduced mortality in patients with sepsis-induced ARDS and those with higher MAP are consistent with the current clinical practice. Additionally, the current study indicates that diuretics may be beneficial in patients with $PaO_2/FiO_2 \leq 150$ mmHg, possibly due to the reduction in EVLW. As EVLW estimates the fluid in pulmonary interstitial and alveolar spaces and is strongly associated with the deterioration of PaO_2/FiO_2, more severe lung injury, and higher mortality (3, 35), decrease in EVLW may be associated with improved survival (36). We postulated that diuretics may have substantially alleviated pulmonary edema in the worse oxygenation subgroup and contributed to better survival.

There are various pathophysiological mechanisms by which loop diuretics may improve the outcomes in patients with ARDS. Diuretics could reduce hydrostatic pressure in the event of alveolar–capillary barrier damage by limiting fluid overload, and may increase colloid osmotic pressure, resulting in reduced pulmonary edema (2, 37). Furthermore, previous studies have suggested that hyper-inflammatory or hypo-inflammatory patients respond differently to randomly assigned fluid management (9). The implementation of liberal or conservative fluid strategy may depend on the inflammatory state, whereas diuretics modify fluid balance and may thus affect the prognosis of certain subtypes.

The present study is the first to explore the effect of loop diuretics use on 28-day mortality in patients with ARDS, using MSCM to account for both time-fixed and time-dependent confounders. The phenotypes derived based on variables

TABLE 3 | Comparisons of the baseline and clinical characteristics between the subtypes.

	Subtype 1	Subtype 2	Subtype 3	*P*
N	89	635	208	
Age, years	46.56 (13.22)	48.34 (16.39)	55.79 (14.23)	<0.001
Female (*n*, %)	33 (37.1)	291 (45.8)	109 (52.4)	0.045
Body mass index	24.23 (6.39)	29.19 (6.91)	30.61 (8.18)	<0.001
APACHE III	110.60 (30.04)	86.76 (28.34)	99.31 (29.53)	<0.001
Primary lung injury (*n*, %)				<0.001
Sepsis	11 (12.4)	136 (21.4)	59 (28.4)	
Trauma	1 (1.1)	68 (10.7)	4 (1.9)	
Aspiration	8 (9.0)	107 (16.9)	31 (14.9)	
Pneumonia	65 (73.0)	272 (42.8)	101 (48.6)	
Other	4 (4.5)	52 (8.2)	13 (6.2)	
Comorbidity (*n*, %)				
Immune suppression	14 (15.7)	30 (4.7)	27 (13.0)	<0.001
Diabetes	12 (13.5)	0 (0.0)	143 (68.8)	<0.001
Hypertension	11 (12.4)	108 (17.0)	102 (49.0)	<0.001
Myocardial infarction	1 (1.1)	12 (1.9)	21 (10.1)	<0.001
Congestive heart failure	1 (1.1)	9 (1.4)	17 (8.2)	<0.001
Chronic pulmonary disease	5 (5.6)	34 (5.4)	22 (10.6)	0.028
Charlson Comorbidity Index	6.00 (6.00, 6.00)	0.00 (0.00, 0.00)	2.00 (2.00, 3.00)	<0.001
Heart rate, bpm	99.79 (19.65)	98.13 (19.29)	93.85 (20.27)	0.011
Respiratory rate, bpm	30.20 (8.30)	27.45 (7.21)	28.03 (7.38)	0.005
CVP, mmHg	10.24 (5.91)	11.64 (4.81)	12.22 (4.85)	0.008
MAP, mmHg	74.24 (11.37)	78.95 (13.80)	77.45 (14.16)	0.008
Vasopressors (*n*, %)	0.27 (0.45)	0.27 (0.45)	0.36 (0.48)	0.071
Fluid balance, ml	2094.21 (2833.58)	1412.33 (3181.05)	2002.85 (3120.95)	0.021
Hemoglobin, g/dl	9.29 (1.55)	10.14 (1.66)	9.49 (1.44)	<0.001
Platelets, $\times 10^{12}$/L	185.94 (100.06)	192.06 (121.98)	194.40 (127.19)	0.862
Creatinine, mg/dl	1.16 (0.73)	1.21 (0.99)	1.56 (1.08)	<0.001
PaO_2/FiO_2, mmHg	144.68 (73.49)	147.50 (62.00)	149.69 (70.95)	0.836
Clinical outcomes				
28-day mortality (*n*, %)	37 (41.6)	92 (14.5)	60 (28.8)	< 0.001
VFDs by day 28, day	21.00 (15.25, 24.00)	21.00 (16.00,24.00)	20.00 (16.00, 24.00)	0.922
RRT by day 90 (*n*, %)	5 (5.7)	34 (5.4)	24 (11.5)	0.008
RRT days, day	31.00 (15.00, 34.00)	16.00 (8.75, 30.00)	12.00 (9.00, 29.50)	0.391

Data are presented as mean (SD), median (IQR), or number (proportion).

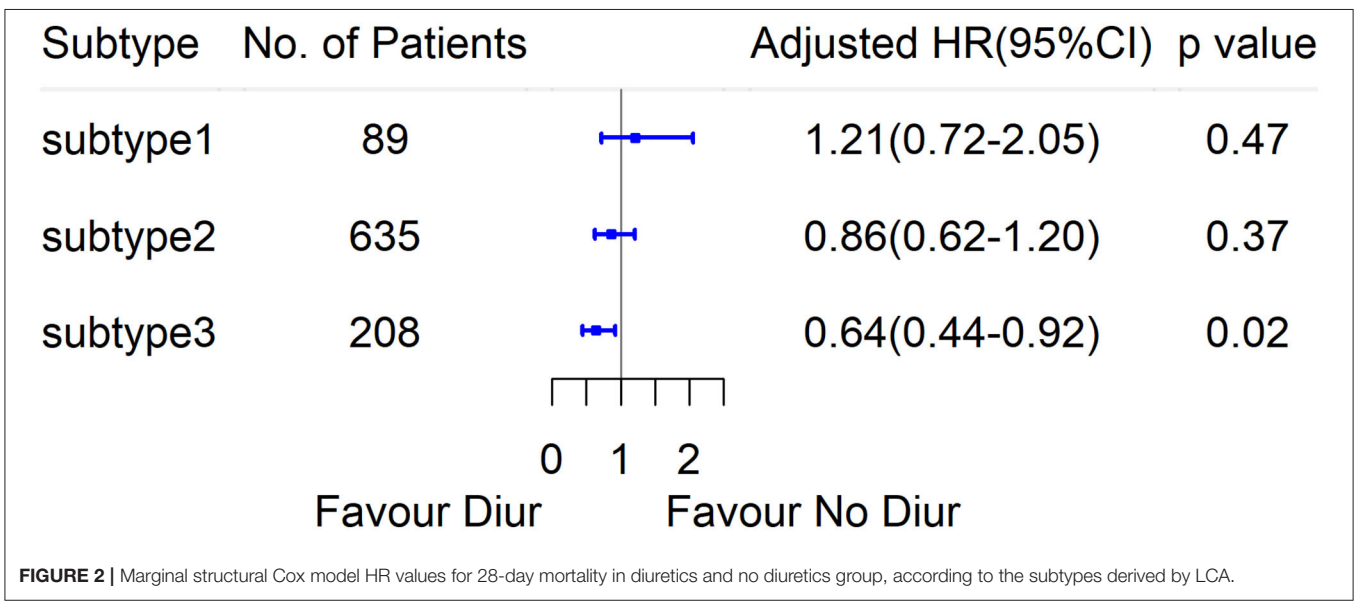


FIGURE 2 | Marginal structural Cox model HR values for 28-day mortality in diuretics and no diuretics group, according to the subtypes derived by LCA.

accessible from medical history and routine laboratory tests may inspire clinicians to implement more precise treatment. Notably, the study had several limitations. First, inflammatory biomarkers were not included in LCA due to limited access to data. Patients were divided into three categories with differences in comorbidities, CVP, and renal function, in accordance with the clinical practice. Another limitation was that we used a dataset over 15 years ago, while this still represents one of the largest randomized clinical trials investigating the effects of fluid strategy on ARDS outcomes and constitutes important evidence-based research of relevance to the guidelines and clinical practice. Last, the retrospective secondary analysis lacked the power to explain the causality. Additional well-designed randomized controlled clinical trials are required.

CONCLUSION

Loop diuretics use was associated with reduced 28-day mortality in the patients with ARDS, after correction for the time-dependent variables. This association was even significant in patients with worse renal function and higher CVP, and in women, patients with sepsis-induced ARDS, and those with lower PaO_2/FiO_2 and higher MAP. The randomized controlled trials are required to validate these results.

DATA AVAILABILITY STATEMENT

The datasets presented in the current study are available in the BioLINCC website (https://biolincc.nhlbi.nih.gov).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Commission of Zhongda Hospital, School of Medicine, Southeast University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

RZ carried out the design, participated in the collection and assembly of data, and drafted the manuscript. HC wrote part of the manuscript. HC, ZG, ML, YY, and HQ participated in the manuscript revision. LL carried out the design, manuscript writing, and final approval of this research. All authors read and approved the final version before submission.

FUNDING

This study was supported by the Clinical Science and Technology Specific Projects of Jiangsu Province (BE2020786 and BE2019749), the National Natural Science Foundation of China (grant number 81870066), and the Natural Science Foundation of Jiangsu Province (BK20171271).

ACKNOWLEDGMENTS

We thank the Intensive Care Unit of Zhongda Hospital, School of Medicine, Southeast University (Nanjing, China) for their helpful and continuous support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.740675/full#supplementary-material

REFERENCES

1. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA.* (2016) 315:788–800. doi: 10.1001/jama.2016.0291
2. Neamu RF, Martin GS. Fluid management in acute respiratory distress syndrome. *Curr Opin Crit Care.* (2013) 19:24–30. doi: 10.1097/MCC.0b013e32835c285b
3. Chew MS, Ihrman L, During J, Bergenzaun L, Ersson A, Undén J, et al. Extravascular lung water index improves the diagnostic accuracy of lung injury in patients with shock. *Crit Care.* (2012) 16:R1. doi: 10.1186/cc10599
4. Jones SL, Martensson J, Glassford NJ, Eastwood GM, Bellomo R. Loop diuretic therapy in the critically ill: a survey. *Crit Care Resusc.* (2015) 17:223–6.
5. Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D, deBoisblanc B, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med.* (2006) 354:2564–75. doi: 10.1056/NEJMoa062200
6. Cinotti R, Lascarrou JB, Azais MA, Colin G, Quenot JP, Mahe PJ, et al. Diuretics decrease fluid balance in patients on invasive mechanical ventilation: the randomized-controlled single blind, IRIHS study. *Crit Care.* (2021) 25:98. doi: 10.1186/s13054-021-03509-5
7. Seitz KP, Caldwell ES, Hough CL. Fluid management in ARDS: an evaluation of current practice and the association between early diuretic use and hospital mortality. *J Intensive Care.* (2020) 8:78. doi: 10.1186/s40560-020-00496-7
8. Bos LD, Schouten LR, van Vught LA, Wiewel MA, Ong DSY, Cremer O, et al. Identification and validation of distinct biological phenotypes in patients with acute respiratory distress syndrome by cluster analysis. *Thorax.* (2017) 72:876–83. doi: 10.1136/thoraxjnl-2016-209719
9. Famous KR, Delucchi K, Ware LB, Kangelaris KN, Liu KD, Thompson BT, et al. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med.* (2017) 195:331–8. doi: 10.1164/rccm.201603-0645OC
10. Bos LDJ, Artigas A, Constantin JM, Hagens LA, Heijnen N, Laffey JG, et al. Precision medicine in acute respiratory distress syndrome: workshop report and recommendations for future research. *EurRespir Rev.* (2021) 30:200317. doi: 10.1183/16000617.0317-2020
11. Libório AB, Barbosa ML, Sá VB, Leite TT. Impact of loop diuretics on critically ill patients with a positive fluid balance. *Anaesthesia.* (2020) 75 (Suppl 1):e134–e42. doi: 10.1111/anae.14908
12. Wheeler AP, Bernard GR, Thompson BT, Schoenfeld D, Wiedemann HP, deBoisblanc B, et al. Pulmonary-artery versus central venous catheter to

guide treatment of acute lung injury. *N Engl J Med.* (2006) 354:2213–24. doi: 10.1056/NEJMoa061895
13. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* (1987) 40:373–83. doi: 10.1016/0021-9681(87)90171-8
14. Murray JF, Matthay MA, Luce JM, Flick MR. An expanded definition of the adult respiratory distress syndrome. *Am Rev Respir Dis.* (1988) 138:720–3. doi: 10.1164/ajrccm/138.3.720
15. Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, et al. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med.* (2018) 6:691–8. doi: 10.1016/S2213-2600(18)30177-2
16. Kim SY. Determining the number of latent classes in single- and multi-phase growth mixture models. *Struct Equation Model.* (2014) 21:263–79. doi: 10.1080/10705511.2014.882690
17. Xie D, Yang W, Jepson C, Roy J, Hsu JY, Shou H, et al. Statistical methods for modeling time-updated exposures in cohort studies of chronic kidney disease. *Clin J Am Soc Nephrol.* (2017) 12:1892–9. doi: 10.2215/CJN.00650117
18. Naimi AI, Moodie EE, Auger N, Kaufman JS. Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology.* (2014) 25:292–9. doi: 10.1097/EDE.0000000000000053
19. van Mourik N, Metske HA, Hofstra JJ, Binnekade JM, Geerts BF, Schultz MJ, et al. Cumulative fluid balance predicts mortality and increases time on mechanical ventilation in ARDS patients: an observational cohort study. *PLoS ONE.* (2019) 14:e0224563. doi: 10.1371/journal.pone.0224563
20. Munshi L, Rubenfeld G, Wunsch H. Adjuvants to mechanical ventilation for acute respiratory distress syndrome. *Intensive Care Med.* (2016) 42:775–8. doi: 10.1007/s00134-016-4327-2
21. Wiedermann CJ. Phases of fluid management and the roles of human albumin solution in perioperative and critically ill patients. *Curr Med Res Opin.* (2020) 36:1961–73. doi: 10.1080/03007995.2020.1840970
22. Silversides JA, Major E, Ferguson AJ, Mann EE, McAuley DF, Marshall JC, et al. Conservative fluid management or deresuscitation for patients with sepsis or acute respiratory distress syndrome following the resuscitation phase of critical illness: a systematic review and meta-analysis. *Intensive Care Med.* (2017) 43:155–70. doi: 10.1007/s00134-016-4573-3
23. Grissom CK, Hirshberg EL, Dickerson JB, Brown SM, Lanspa MJ, Liu KD, et al. Fluid management with a simplified conservative protocol for the acute respiratory distress syndrome*. *Crit Care Med.* (2015) 43:288–95. doi: 10.1097/CCM.0000000000000715
24. Cordemans C, De Laet I, Van Regenmortel N, Schoonheydt K, Dits H, Martin G, et al. Aiming for a negative fluid balance in patients with acute lung injury and increased intra-abdominal pressure: a pilot study looking at the effects of PAL-treatment. *Ann Intensive Care.* (2012) 2 (Suppl 1):S15. doi: 10.1186/2110-5820-2-S1-S15
25. Martin GS, Gattinoni L, Chiumello D. Fluid administration and monitoring in ARDS: which management? *Intensive Care Med.* (2020) 46:2252–64. doi: 10.1007/s00134-020-06310-0
26. Zinter MS, Spicer AC, Liu KD, Orwoll BE, Alkhouli MF, Brakeman PR, et al. Positive cumulative fluid balance is associated with mortality in pediatric acute respiratory distress syndrome in the setting of acute kidney injury. *Pediatr Crit Care Med.* (2019) 20:323–31. doi: 10.1097/PCC.0000000000001845
27. Joannidis M, Klein SJ, Ostermann M. 10 myths about frusemide. *Intensive Care Med.* (2019) 45:545–8. doi: 10.1007/s00134-018-5502-4
28. Zhao GJ, Xu C, Ying JC, Lü WB, Hong GL, Li MF, et al. Association between furosemide administration and outcomes in critically ill patients with acute kidney injury. *Crit Care.* (2020) 24:75. doi: 10.1186/s13054-020-2798-6
29. Shen Y, Zhang W, Shen Y. Early diuretic use and mortality in critically ill patients with vasopressor support: a propensity score-matching analysis. *Crit Care.* (2019) 23:9. doi: 10.1186/s13054-019-2309-9
30. Mullens W, Damman K, Harjola VP, Mebazaa A, Brunner-La Rocca HP, Martens P, et al. The use of diuretics in heart failure with congestion—a position statement from the Heart Failure Association of the European Society of Cardiology. *Eur J Heart Failure.* (2019) 21:137–55. doi: 10.1002/ejhf.1369
31. Unger T, Borghi C, Charchar F, Khan NA, Poulter NR, Prabhakaran D, et al. 2020 international society of hypertension global hypertension practice guidelines. *Hypertension.* (2020) 75:1334–57. doi: 10.1161/HYPERTENSIONAHA.120.15026
32. Rhodes A, Evans LE, Alhazzani W, Levy MM, Antonelli M, Ferrer R, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. *Intensive Care Med.* (2017) 43:304–77. doi: 10.1007/s00134-017-4683-6
33. Acheampong A, Vincent JL. A positive fluid balance is an independent prognostic factor in patients with sepsis. *Crit Care.* (2015) 19:251. doi: 10.1186/s13054-015-0970-1
34. Murphy CV, Schramm GE, Doherty JA, Reichley RM, Gajic O, Afessa B, et al. The importance of fluid management in acute lung injury secondary to septic shock. *Chest.* (2009) 136:102–9. doi: 10.1378/chest.08-2706
35. Berkowitz DM, Danai PA, Eaton S, Moss M, Martin GS. Accurate characterization of extravascular lung water in acute respiratory distress syndrome. *Crit Care Med.* (2008) 36:1803–9. doi: 10.1097/CCM.0b013e3181743eeb
36. Tagami T, Nakamura T, Kushimoto S, Tosa R, Watanabe A, Kaneko T, et al. Early-phase changes of extravascular lung water index as a prognostic indicator in acute respiratory distress syndrome patients. *Ann Intensive Care.* (2014) 4:27. doi: 10.1186/s13613-014-0027-7
37. Martin GS, Mangialardi RJ, Wheeler AP, Dupont WD, Morris JA, Bernard GR. Albumin and furosemide therapy in hypoproteinemic patients with acute lung injury. *Crit Care Med.* (2002) 30:2175–82. doi: 10.1097/00003246-200210000-00001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 23 September 2021
doi: 10.3389/fmed.2021.707437

Deep Learning Assisted Detection of Abdominal Free Fluid in Morison's Pouch During Focused Assessment With Sonography in Trauma

Chi-Yung Cheng[1,2], *I-Min Chiu*[1,2]*, *Ming-Ya Hsu*[2], *Hsiu-Yung Pan*[1], *Chih-Min Tsai*[3] and *Chun-Hung Richard Lin*[2]*

[1] Department of Emergency Medicine, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung, Taiwan, [2] Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan, [3] Department of Pediatrics, Kaohsiung Chang Gung Memorial Hospital, Chang Gung University College of Medicine, Kaohsiung, Taiwan

OPEN ACCESS

Edited by:
Qinghe Meng,
Upstate Medical University,
United States

Reviewed by:
Yan Ren Lin,
Changhua Christian Hospital, Taiwan
Yichun Wang,
The Third Affiliated Hospital of
Gaungzhou Medical University, China

***Correspondence:**
I-Min Chiu
ray1985@cgmh.org.tw
Chun-Hung Richard Lin
lin@cse.nsysu.edu.tw

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: 10 May 2021
Accepted: 25 August 2021
Published: 23 September 2021

Citation:
Cheng C-Y, Chiu I-M, Hsu M-Y, Pan H-Y, Tsai C-M and Lin C-HR (2021) Deep Learning Assisted Detection of Abdominal Free Fluid in Morison's Pouch During Focused Assessment With Sonography in Trauma. Front. Med. 8:707437. doi: 10.3389/fmed.2021.707437

Background: The use of focused assessment with sonography in trauma (FAST) enables clinicians to rapidly screen for injury at the bedsides of patients. Pre-hospital FAST improves diagnostic accuracy and streamlines patient care, leading to dispositions to appropriate treatment centers. In this study, we determine the accuracy of artificial intelligence model-assisted free-fluid detection in FAST examinations, and subsequently establish an automated feedback system, which can help inexperienced sonographers improve their interpretation ability and image acquisition skills.

Methods: This is a single-center study of patients admitted to the emergency room from January 2020 to March 2021. We collected 324 patient records for the training model, 36 patient records for validation, and another 36 patient records for testing. We balanced positive and negative Morison's pouch free-fluid detection groups in a 1:1 ratio. The deep learning (DL) model Residual Networks 50-Version 2 (ResNet50-V2) was used for training and validation.

Results: The accuracy, sensitivity, and specificity of the model performance for ascites prediction were 0.961, 0.976, and 0.947, respectively, in the validation set and 0.967, 0.985, and 0.913, respectively, in the test set. Regarding feedback prediction, the model correctly classified qualified and non-qualified images with an accuracy of 0.941 in both the validation and test sets.

Conclusions: The DL algorithm in ResNet50-V2 is able to detect free fluid in Morison's pouch with high accuracy. The automated feedback and instruction system could help inexperienced sonographers improve their interpretation ability and image acquisition skills.

Keywords: deep learning, FAST, Morison pouch, ascites, trauma, hemoperitoneum

INTRODUCTION

Traumatic injury remains the leading cause of death among individuals younger than 45 years of age (1), with over 210,000 deaths per year in the last 5 years in the United States (2). A substantial proportion of such patients suffered from blunt abdominal trauma (3). Computed tomography had been the gold standard for diagnosing intra-abdominal or thoracic injuries. However, time delays and transportation out of the emergency department (ED) hinders the evaluation of hemodynamically unstable patients. The use of focused assessment with sonography in trauma (FAST) enabled clinicians to rapidly screen for injury at the bedsides of patients. Recent studies have shown that FAST plays a key role in trauma detection, changing the subsequent management of an appreciable number of patients. In addition, pre-hospital FAST improves diagnostic accuracy and streamlines patient care, leading to dispositions to appropriate treatment centers. It was previously demonstrated that after performing pre-hospital FAST, pre-hospital therapy and management can be altered for 30% of the patients, and patient disposition can occur in 22% of the cases being admitted to the ED (4–6). A major limitation of ultrasound is that it is operator-dependent, with training, experience, and inter-operator variability playing an important role (7).

Artificial intelligence (AI), a subfield of computer science, helps create systems that perform tasks in medicine, medically oriented human biology, and healthcare improvements. Machine learning (ML), a subfield of AI, outperforms traditional approaches to various diseases and clinical conditions, including diagnosis, quality of patient care, and prognosis of a disease (8–14). Compared with traditional approaches, ML procedures may have the ability to interact with non-linear and high-order effects in variable parameters. Owing to the nature of operator-dependent imaging modality in ultrasound, developing deep learning (DL) models that assess image quality and provide feedback to sonographers was considered to provide ultrasound with more intelligence. AI-assisted ultrasound is expected to minimize operator-dependent imaging modality, altering medical therapy and patient disposition in critical care units and pre-hospital care.

Owing to the amorphous nature of intra-abdominal free fluid, AI-assisted ascites detection remains a challenge. Our study aims to determine the accuracy of AI model-assisted free-fluid detection during FAST examination and subsequently establish an automated feedback system, which can help inexperienced sonographers improve their interpretation ability and image acquisition skills. Moreover, AI model-assisted real-time ultrasound could enhance the diagnostic performance of FAST when used by paramedics or during an emergency.

MATERIALS AND METHODS

Study Setting and Variables

This is a single-center study of patients admitted to the ED at Kaohsiung Chang Gung Memorial Hospital, Taiwan. Abdominal ultrasound clips were taken for a variety of clinical conditions and saved in the emergency ultrasound image archive in an MPEG-4 format. These clips were taken by 10 certified attending emergency physicians using a time-motion ultrasound machine with a 5–2 MHz curved mechanical sector transducer. The study was approved by the IRB committee of the hospital (IRB number: 202001766B0C601)

For the training set, all patients aged >18 years who underwent abdominal ultrasounds in the ED from January 2020 to October 2020 were included. Because of the study's retrospective nature in this study period, informed consent from the subjects was not required. Ultrasound examinations were retrospectively reviewed and retrieved from the image database in the ED during the study period. We only retrieved examinations performed on the right upper abdominal quadrant for Morison's pouch scanning. Morison's pouch is the space that separates the liver from the right kidney, considered the lowest intra-abdominal area for detecting free fluid in the supine position.

For the validation and test sets, patients aged >18 years who underwent FAST study in the ED from November 2020 to March 2021 were included. Informed consent was obtained from all subjects involved during this study period prior to beginning the abdominal ultrasound examinations.

To compare the ascites interpretation between emergency medicine (EM) residents and model performance, 10 registered EM residents were recruited for the trial from hospital personnel. Each EM resident had received at least a 1-year training course in the ED. The result for the EM residents' ultrasound finding interpretation of the test set was compared to that of the DL model in terms of accuracy, sensitivity, and specificity.

Data Pre-processing and Labeling

All collected ultrasound videos were first converted to still images at a rate of 10 frames per second with an initial size of 800 × 600 pixels. Subsequently, each image was reviewed by 4 ultrasound instructors in the hospital to determine whether it was positive or negative for free-fluid detection.

A feedback labeling for the standard Morison's pouch view was subsequently added during the image review process to implement the assisting system that enabled the operator to distinguish whether the current image was qualified to detect the Morison's pouch fluid. The qualified view was defined as the area between the liver and kidney, caudal edge of the liver, or right paracolic gutter area. The image was classified as a non-qualified view if less than one-third of the right kidney was observable.

After the review process, each image was classified into one of four classes: positive/qualified view, positive/non-qualified view, negative/qualified view, and negative/non-qualified view. All images were labeled by four qualified ultrasound instructors who were certified attending emergency physicians in Taiwan, and at least 3 out of 4 instructors had to agree on the classification of each image.

Deep Learning Training Process

To avoid overfitting the model, all labeled images were cropped to a size of 400 × 400 pixels to remove unnecessary information such as the background grid and knobology settings information. Other image augmentation techniques, namely,

random rotation, random zoom, and horizontal flip, were applied to the training set.

In this study, we used the DL model Residual Networks 50-Version 2 (ResNet50-V2) for training and validation (15). With limited clinical data, we obtained model weights of ResNet50-V2 from the ImageNet database (16) as a pre-trained model and performed transfer learning through a fine-tuning process during training. During transfer learning, we froze the model weights in convolution layers 1 to 3 and updated the weights in convolution layers 4 and 5 and in the fully connected neural network on the top during training. All the aforementioned deep neural networks were developed using Python 3.8 and TensorFlow version 2.4.1.

Statistical Analysis

We balanced positive and negative free-fluid detection groups with a 1:1 ratio during model training, validation, and testing to prevent deviations due to imbalanced data. We collected 162 patient records with positive free-fluid detection and randomly extracted 162 patient records with negative findings from the image database for the training model. We also collected 18 patient records each from the positive and negative groups for validation and another 18 patient records each for testing. Each patient record contained 15–35 second long ultrasound videos. During the DL training process, we tried to optimize the model to achieve the best performance in the validation set. The final performance was evaluated in the test set as an external validation.

The performance of the model for fluid detection was assessed in terms of accuracy, sensitivity, and specificity. The performance of the feedback was evaluated based on accuracy. In addition, from a clinical perspective, sensitivity is considered more relevant than specificity because it shows how accurately the intra-abdominal free fluid was identified; thus, the level of sensitivity was prioritized over specificity during training and validation.

For a real-time ultrasound assisting system, predicting a class for every input image might be difficult. Alternatively, generating a prediction every 10 frames is more feasible in clinical practice. Consequently, we also considered another evaluation strategy for model performance on ascites prediction in both validation and test sets, which employed a majority-voting scheme for consecutive images in a 1-s window, where the majority of the image class predictions were taken in that specific time frame.

In this study, continuous variables were presented as mean ± standard deviation. Dichotomous data were presented as numbers (percentages). Categorical variables were analyzed using the chi-square test, and continuous variables were analyzed using the independent sample *t*-test. All statistical analyses were performed using SPSS 26.

RESULTS

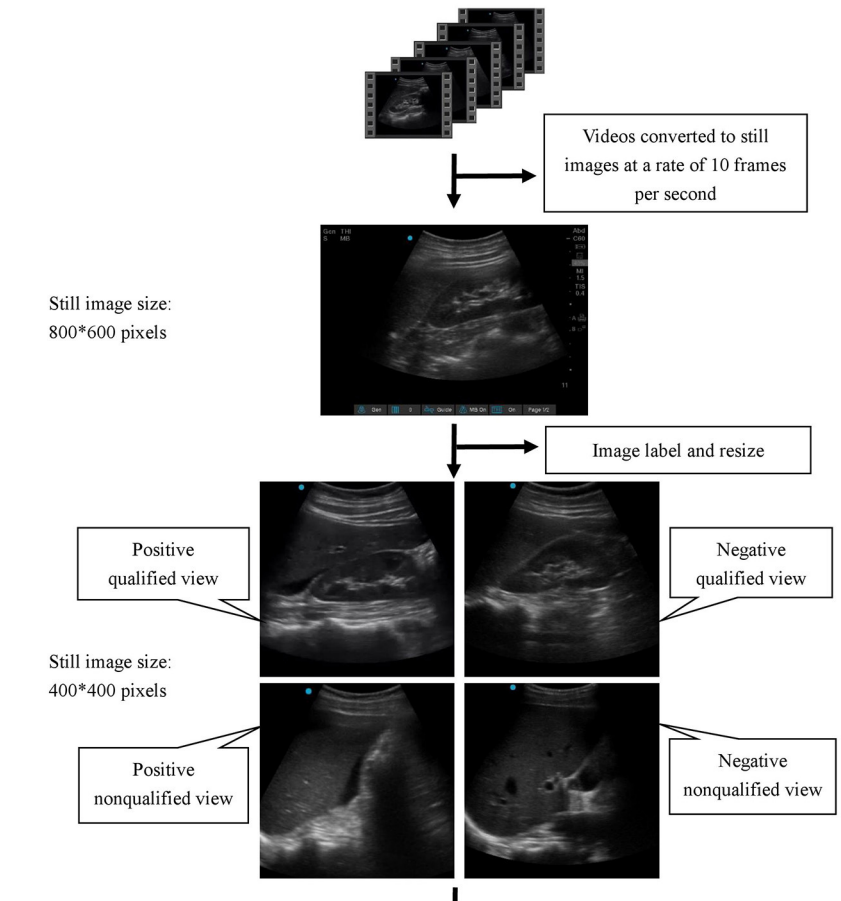
The demographic characteristics of the patients included in the study are listed in **Table 1**. There was no significant difference in age, sex, body mass index, and underlying diseases between the training set, validation set, and test set. For training, validation, and testing, 10,794, 6,118, and 5,456 images were, respectively, included. In the validation set, there were 3,121 negative and 2,997 positive ascites images. Among them, 3,750 images were labeled as qualified images. In the 2,997 positive ascites images, there were 1,488 frames with mild ascites, 774 with moderate ascites, and 735 with massive ascites. In the test set, there were 2,780 positive and 2,676 negative ascites images. Among them, 3,337 images were labeled as qualified images. In the 2,676 positive ascites images, there were 1,332 frames with mild ascites, 682 with moderate ascites, and 662 with massive ascites. **Figure 1** depicts a flowchart of the dataset construction process.

To prevent the model from overfitting during training, we added a batch normalization layer and a dropout layer on top of the last fully connected neural layer of the final developed model. We used the Adam algorithm as an optimizer with an initial learning rate of 2×10^{-6} and adjusted the class weight to favor prediction over positive ascites for clinical priority purposes. The model was trained for 100 epochs, and the best model weights

TABLE 1 | Demographic data of patients.

	Training set (*n* = 324)	Validation set (*n* = 36)	Test set (*n* = 36)	*P* value
Demographic characteristics				
Age, years, mean ± SD	59.7 ± 17.2	61.1 ± 15.9	60.5 ± 19.2	0.889
Sex, male, *n* (%)	174 (53.7)	19 (52.8)	20 (55.6)	0.981
BH, mean ± SD	162.2 ± 8.8	161.8 ± 9.8	163.4 ± 8.5	0.731
BW, mean ± SD	60.9 ± 13.1	62.6 ± 14.2	62.3 ± 12.0	0.661
BMI (kg/m^2), mean ± SD	23.3 ± 4.4	23.6 ± 4.3	23.4 ± 4.1	0.968
Underlying disease				
Heart failure, *n* (%)	19 (5.9)	3 (8.3)	2 (5.6)	0.779
Chronic kidney disease, *n* (%)	57 (17.6)	6 (16.7)	5 (13.9)	0.906
Liver cirrhosis, *n* (%)	40 (12.3)	5 (13.9)	6 (16.7)	0.840
Malignancy, *n* (%)	54 (16.7)	4 (11.1)	8 (13.9)	0.699

BH, body height; BW, body weight; BMI, body mass index.



Training Set
Number of patients
Positive: negative = 162: 162
Number of still images

	Positive	Negative
Qualified	5263	5325
Nonqualified	5095	5111

Validation Set
Number of patients
Positive: negative = 18: 18
Number of still images

	Positive	Negative
Qualified	1829	1925
Nonqualified	1168	1196

Test Set
Number of patients
Positive: negative = 18: 18
Number of still images

	Positive	Negative
Qualified	1846	1891
Nonqualified	830	889

FIGURE 1 | Flowchart of dataset construction.

during training were saved and evaluated in the validation and test datasets.

The confusion matrix for the four class prediction results in the validation and test sets are shown in **Figure 2**. The accuracy, sensitivity, and specificity of the model performance for ascites prediction were 0.961, 0.976, and 0.947, respectively, in the validation set and 0.967, 0.985, and 0.913, respectively, in the test set (**Table 2**). For ascites prediction in the EM resident group, the accuracy, sensitivity, and specificity were 0.966, 0.989, and 0.943, respectively (**Table 3**). The result for human interpretation was not significantly different compared with the DL model ($p = 0.570$). Regarding feedback prediction, the model correctly classified qualified and non-qualified images with an accuracy of 0.941 in both the validation and test sets (**Table 2**).

By using the aforementioned majority-voting scheme for evaluation, the model was able to identify every ascites clip in both the validation and test sets, while it misclassified only two negative ascites frames: one into the positive class in the validation set and one into the positive class in the test set. The results of all the prediction performances are provided in **Table 2**. The accuracy, sensitivity, and specificity of the resident physician vs. model performance were, respectively, 0.986 vs. 0.998, 1 vs. 1, and 0.972 vs. 0.996 ($p = 0.001$).

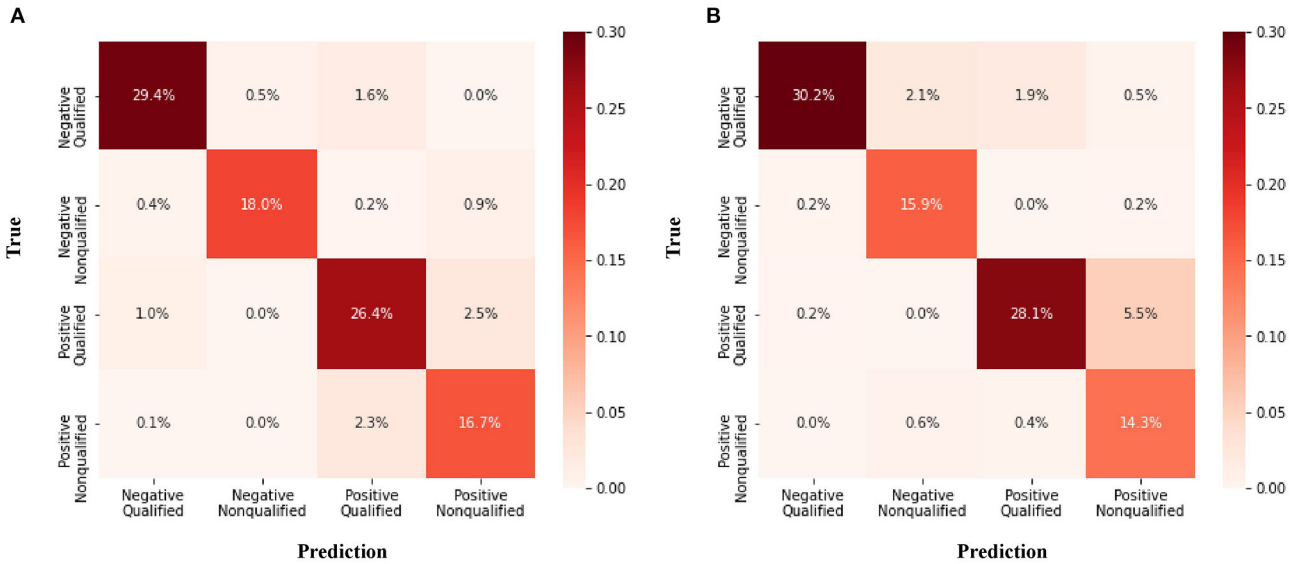


FIGURE 2 | Confusion matrix for the four class prediction results. **(A)** Validation set **(B)** Test set.

TABLE 2 | Model performance for ascites and image location feedback prediction.

	Validation set	Test set
By frame	**(*n* = 6,118)**	**(*n* = 5,456)**
Ascites prediction		
Accuracy	0.961	0.967
Sensitivity	0.976	0.985
PPV	0.947	0.949
Specificity	0.947	0.913
Feedback prediction		
Accuracy	0.941	0.941
By 1-s majority voting	**(*n* = 309)**	**(*n* = 500)**
Ascites prediction		
Accuracy	0.997	0.998
PPV	0.993	0.996
Sensitivity	1	1
Specificity	0.994	0.996

PPV, Positive predictive value.

DISCUSSION

In our study, we detected abdominal free fluid and predicted the location of the Morison pouch using only a single frame of the ultrasound image. ResNet50-V2 was able to detect the abdominal free fluid of the Morison pouch with accuracy, sensitivity, and specificity values of 96.1, 97.6, and 94.7%, respectively, in the validation set and 96.7, 98.5, and 91.3%, respectively, in the test set. The result of ResNet50-V2 performance was non-inferior to the EM resident interpretation. By using the majority-voting scheme for consecutive images in a 1-s window, the DL model was able to reach 100% sensitivity, and the specificity was significantly better than the EM resident interpretation. Previous studies demonstrated that FAST examinations with human interpretation for intra-peritoneal free-fluid detection have a sensitivity ranging from 61.3 to 100% and specificity ranging from 94 to 100% for blunt abdominal trauma (17–20). The sensitivity of the ultrasound examination (28–100%) was considered insufficient for it to be used alone in determining operative intervention for penetrating torso trauma (21–23). Although initially developed for the evaluation of trauma patients, FAST examination can also be used in non-trauma patients to narrow down differential diagnoses, change patient disposition, expedite consultation, avoid unnecessary procedures, and alter imaging needs (24). However, studies of sensitivity and specificity are limited owing to the large variety of etiologies in non-trauma patients.

Applications of DL in medical ultrasound analysis include anatomical applications, diagnosis tasks (classification, segmentation, detection), and clinical tasks (computer-aided diagnosis, biometric measurements, image-guided interventions) (8). DL models have been used to detect different anatomical structures of human organs in medical analyses, including the brain (9), heart (10), thyroid (11), breast (12), liver (13), and prostate (14). Recognized as one of the most popular deep architectures, the convolution neural network (CNN) has been applied to various tasks, such as image classification, object detection, and target segmentation (9, 25). Our study shows that a computer program developed incorporating ResNet50-V2 could aid in the detection of free fluids in FAST examination, with the results obtained on par with the interpretations of a medical doctor.

In various ultrasound protocols, performance plateaus occur at different points for image interpretation and quality (26). In addition, physicians acquire the ability to interpret FAST images quicker than they acquire the technical skills required

TABLE 3 | Comparison of resident doctor and model performance for ascites interpretation.

	Resident physician (n=10)	ResNet50-V2 model	*P* value
By frame	**(*n* = 5,456)**		
Accuracy	0.966	0.967	0.570
Sensitivity	0.989	0.985	
Specificity	0.943	0.913	
By 1-s majority voting	**(*n* = 500)**		
Accuracy	0.986	0.998	0.033*
Sensitivity	1	1	
Specificity	0.972	0.996	

*$p < 0.05$.

to perform the examination (27). Focusing on the acquisition of images for FAST examination, Jang et al. (27) found that the ultrasound technique continued to improve even after 75 examinations. Blehar et al. demonstrated that the learning curve in image quality for FAST examination improved even after 200 examinations (26). In our model, the accuracy of locating Morison's pouch reached 94.05 and 91.35% in the validation and test sets, respectively. The automated feedback and instruction system is believed to assist inexperienced sonographers improve their interpretation ability and image acquisition skills.

Free-fluid detection by ultrasound could be used in both trauma and non-trauma patients and could have a broad impact on patient care across a wide range of medical settings (28). In addition, AI model-assisted real-time ultrasound could enhance the diagnostic performance of FAST when used by paramedics or during an emergency. Using FAST examination in the pre-hospital stage can significantly improve the outcomes of blunt abdominal trauma (5). However, to obtain the benefits of AI model-assisted FAST in-patient care, additional large-scale studies should assess the performance of free-fluid detection for different trauma and non-trauma etiologies.

The limitations in our study are discussed below. First, we trained using only perihepatic views in our study. In certain conditions, a single Morison's pouch view was employed because the right upper quadrant was considered the primary area where free fluid is initially seen and the most sensitive for free-fluid assessment (29, 30). However, a multiple-view FAST examination was recommended to increase sensitivity (31). Second, we tested our model with a single-frame ultrasound image and 1-s majority voting, which may have led to variable sensitivity and specificity. The purpose of free-fluid detection should be for identification using serial video clips. Third, the development of automated feedback and instructions by AI model-assisted free-fluid detection remains a challenge. For the classification of qualified or non-qualified view, the binary classifications in our study potentially caused a loss of continuous information. The feasibility of a real-time AI model-assisted system should be tested in future studies, including ultrasound acquisition software, real-time feedback speed, and indication for proper probe location. Finally, the performance of free-fluid detection for different etiologies should also be considered. Future studies should also include a separate assessment of the test performance for different free fluid etiologies, such as blunt trauma, penetrating trauma, and various non-trauma etiologies.

AI model-assisted real-time ultrasound could enhance the diagnostic performance of FAST when used by paramedics or during an emergency. The DL algorithm with ResNet50-V2, used in our study, was able to detect free fluid in Morison's pouch with accuracies reaching 94.05 and 91.35% in the validation and test sets, respectively. By using the majority-voting scheme for consecutive images in a 1-s window, the DL model was able to reach 100% sensitivity, and the specificity was significantly better than the EM resident interpretation. In the future, AI-assisted ultrasound will minimize operator-dependent imaging modality and alter medical therapy and disposition in critical care units and prehospital care.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Chang Gung Medical Foundation (protocol code: 202001766B0C601). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

C-YC conceived the research, performed the analyses, and wrote the manuscript. H-YP and C-MT contributed to data collection and measurements. M-YH was involved mainly in data analysis and quality management. I-MC and C-HL provided overall supervision, edited the manuscript, and undertook the responsibility of submitting the manuscript for publication. All authors read and approved the final manuscript.

FUNDING

This research was funded by the Ministry of Science and Technology research grant in Taiwan (MOST 110-2314-B-182A-009).

ACKNOWLEDGMENTS

We gratefully acknowledge all persons who helped in this research. This study was supported by the research deputy of Kaohsiung Chang Gung Memorial Hospital.

REFERENCES

1. Melniker LA, Leibner E, McKenney MG, Lopez P, Briggs WM, Mancuso CA. Randomized controlled clinical trial of point-of-care, limited ultrasonography for trauma in the emergency department: the first sonography outcomes assessment program trial. *Ann Emerg Med*. (2006) 48:227–35. doi: 10.1016/j.annemergmed.2006.01.008
2. Centers for Disease Control and Prevention. *National Center for Injury Prevention and Control*. Available online at: https://wisqars-viz.cdc.gov:8006/explore-data/home. (accessed June 27, 2021).
3. Sobrino J, Shafi S. Timing and causes of death after injuries. *Proc (Bayl Univ Med Cent)*. (2013) 26:120–3. doi: 10.1080/08998280.2013.11928934
4. Jørgensen H, Jensen CH, Dirks J. Does prehospital ultrasound improve treatment of the trauma patient? A systematic review. *Eur J Emerg Med*. (2010) 17:249–53. doi: 10.1097/MEJ.0b013e328336adce
5. Walcher F, Weinlich M, Conrad G, Schweigkofler U, Breitkreutz R, Kirschning T, et al. Prehospital ultrasound imaging improves management of abdominal trauma. *Br J Surg*. (2006) 93:238–42. doi: 10.1002/bjs.5213
6. Mazur SM, Pearce A, Alfred S, Goudie A, Sharley P. The F.A.S.T.E.R. trial. Focused assessment by sonography in trauma during emergency retrieval: a feasibility study. *Injury*. (2008) 39:512–8. doi: 10.1016/j.injury.2007.11.010
7. Limchareon S, Jaidee W. Physician-performed Focused Ultrasound: an update on its role and performance. *J Med Ultrasound*. (2015) 23: 67–70. doi: 10.1016/j.jmu.2015.02.006
8. Liu S, Wang Y, Yang X, Lei B, Liu L, Li S, et al. Deep learning in medical ultrasound analysis: a review. *Engineering*. (2019) 5:261–75. doi: 10.1016/j.eng.2018.11.020
9. Cheng PM, Malhi HS. Transfer learning with convolutional neural networks for classification of abdominal ultrasound images. *J Digit Imaging*. (2017) 30:234–43. doi: 10.1007/s10278-016-9929-2
10. Pereira F, Bueno A, Rodriguez A, Perrin D, Marx G, Cardinale M, et al. Automated detection of coarctation of aorta in neonates from two-dimensional echocardiograms. *J Med Imaging*. (2017) 4:014502. doi: 10.1117/1.JMI.4.1.014502
11. Ma J, Wu F, Jiang T, Zhu J, Kong D. Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images. *Med Phys*. (2017) 44:1678–91. doi: 10.1002/mp.12134
12. Yap MH, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE J Biomed Health Inform*. (2018) 22:1218–26. doi: 10.1109/JBHI.2017.2731873
13. Wu K, Chen X, Ding M. Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound. *Optik*. (2014) 125:4057–63. doi: 10.1016/j.ijleo.2014.01.114
14. Azizi S, Bayat S, Yan P, Tahmasebi A, Nir G, Kwak JT, et al. Detection and grading of prostate cancer using temporal enhanced ultrasound: combining deep neural networks and tissue mimicking simulations. *Int J Comput Assist Radiol Surg*. (2017) 12:1293–305. doi: 10.1007/s11548-017-1627-0
15. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: Caesars Palace (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
16. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL: IEEE (2009). p. 248–55. doi: 10.1109/CVPR.2009.5206848
17. Kim CH, Shin SD, Song KJ, Park CB. Diagnostic accuracy of focused assessment with sonography for trauma (FAST) examinations performed by emergency medical technicians. *Prehosp Emerg Care*. (2012) 16:400–6. doi: 10.3109/10903127.2012.664242
18. Brooks A, Davies B, Smethhurst M, Connolly J. Prospective evaluation of non-radiologist performed emergency abdominal ultrasound for haemoperitoneum. *Emerg Med J*. (2004) 21:e5. doi: 10.1136/emj.2003.006932
19. Rozycki GS, Ochsner MG, Schmidt JA, Frankel HL, Davis TP, Wang D, et al. A prospective study of surgeon-performed ultrasound as the primary adjuvant modality for injured patient assessment. *J Trauma*. (1995) 39:492–8. doi: 10.1097/00005373-199509000-00016
20. Soundappan SV, Holland AJ, Cass DT, Lam A. Diagnostic accuracy of surgeon-performed focused abdominal sonography (FAST) in blunt paediatric trauma. *Injury*. (2005) 36:970–5. doi: 10.1016/j.injury.2005.02.026
21. Biffl WL, Kaups KL, Cothren CC, Brasel KJ, Dicker RA, Bullard MK, et al. Management of patients with anterior abdominal stab wounds: a Western Trauma Association multicenter trial. *J Trauma*. (2009) 66:1294–301. doi: 10.1097/TA.0b013e31819dc688
22. Boulanger BR, Kearney PA, Tsuei B, Ochoa JB. The routine use of sonography in penetrating torso injury is beneficial. *J Trauma*. (2001) 51:320–5. doi: 10.1097/00005373-200108000-00015
23. Soffer D, McKenney MG, Cohn S, Garcia-Roca R, Namias N, Schulman C, et al. A prospective evaluation of ultrasonography for the diagnosis of penetrating torso injury. *J Trauma*. (2004) 56:953–7. doi: 10.1097/01.TA.0000127806.39852.4E
24. Maitra S, Jarman R, Halford N, Richards S. When FAST is a FAFF: is FAST scanning useful in non-trauma patients? *Ultrasound*. (2008) 16:165–8. doi: 10.1179/174313408X322750
25. Hafiane A, Vieyres P, Delbos A. Deep learning with spatiotemporal consistency for nerve segmentation in ultrasound images. (2017). arXiv:1706.05870.
26. Blehar DJ, Barton B, Gaspari RJ. Learning curves in emergency ultrasound education. *Acad Emerg Med*. (2015) 22:574–82. doi: 10.1111/acem.12653
27. Jang T, Kryder G, Sineff S, Naunheim R, Aubin C, Kaji AH. The technical errors of physicians learning to perform focused assessment with sonography in trauma. *Acad Emerg Med*. (2012) 19:98–101. doi: 10.1111/j.1553-2712.2011.01242.x
28. Polk JD, Fallon WF Jr, Kovach B, Mancuso C, Stephens M, Malangoni MA. The "Airmedical F.A.S.T." for trauma patients–the initial report of a novel application for sonography. *Aviat Space Environ Med*. (2001) 72:432–6.
29. Von Kuenssberg Jehle D, Stiller G, Wagner D. Sensitivity in detecting free intraperitoneal fluid with the pelvic views of the FAST exam. *Am J Emerg Med*. (2003) 21:476–8. doi: 10.1016/S0735-6757(03)00162-1
30. Lobo V, Hunter-Behrend M, Cullnan E, Higbee R, Phillips C, Williams S, et al. Caudal edge of the liver in the Right Upper Quadrant (RUQ) view is the most sensitive area for free fluid on the FAST exam. *West J Emerg Med*. (2017) 18:270–80. doi: 10.5811/westjem.2016.11.30435
31. Rose JS, Bair AE, Mandavia D, Kinser DJ. The UHP ultrasound protocol: a novel ultrasound approach to the empiric evaluation of the undifferentiated hypotensive patient. *Am J Emerg Med*. (2001) 19:299–302. doi: 10.1053/ajem.2001.24481

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 05 November 2021
doi: 10.3389/fmed.2021.775511

Revising Host Phenotypes of Sepsis Using Microbiology

Huiying Zhao[1,2,3], Jason N. Kennedy[2,3], Shu Wang[4], Emily B. Brant[2,3], Gordon R. Bernard[5], Kimberley DeMerle[3], Chung-Chou H. Chang[6], Derek C. Angus[2,3] and Christopher W. Seymour[2,3,7]*

[1] Department of Critical Care Medicine, Peking University People's Hospital, Beijing, China, [2] Department of Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, United States, [3] Clinical Research, Investigation, and Systems Modeling of Acute Illness (CRISMA) Center, Pittsburgh, PA, United States, [4] Department of Biostatistics, University of Florida, Gainesville, FL, United States, [5] Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, United States, [6] Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA, United States, [7] Department of Emergency Medicine, University of Pittsburgh, Pittsburgh, PA, United States

Background: There is wide heterogeneity in sepsis in causative pathogens, host response, organ dysfunction, and outcomes. Clinical and biologic phenotypes of sepsis are proposed, but the role of pathogen data on sepsis classification is unknown.

Methods: We conducted a secondary analysis of the Recombinant Human Activated Protein C (rhAPC) Worldwide Evaluation in Severe Sepsis (PROWESS) Study. We used latent class analysis (LCA) to identify sepsis phenotypes using, (i) only clinical variables ("host model") and, (ii) combining clinical with microbiology variables (e.g., site of infection, culture-derived pathogen type, and anti-microbial resistance characteristics, "host-pathogen model"). We describe clinical characteristics, serum biomarkers, and outcomes of host and host-pathogen models. We tested the treatment effects of rhAPC by phenotype using Kaplan-Meier curves.

Results: Among 1,690 subjects with severe sepsis, latent class modeling derived a 4-class host model and a 4-class host-pathogen model. In the host model, alpha type (N = 327, 19%) was younger and had less shock; beta type (N=518, 31%) was older with more comorbidities; gamma type (N = 532, 32%) had more pulmonary dysfunction; delta type (N = 313, 19%) had more liver, renal and hematologic dysfunction and shock. After the addition of microbiologic variables, 772 (46%) patients changed phenotype membership, and the median probability of phenotype membership increased from 0.95 to 0.97 ($P < 0.01$). When microbiology data were added, the contribution of individual variables to phenotypes showed greater change for beta and gamma types. In beta type, the proportion of abdominal infections (from 20 to 40%) increased, while gamma type patients had an increased rate of lung infections (from 50 to 78%) with worsening pulmonary function. Markers of coagulation such as d-dimer and plasminogen activator inhibitor (PAI)-1 were greater in the beta type and lower in the gamma type. The 28 day mortality was significantly different for individual phenotypes in host and host-pathogen models (both $P<0.01$). The treatment effect of rhAPC obviously changed in gamma type when microbiology data were added (P-values of log rank test changed from 0.047 to 0.780).

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Yifan Yang,
Transwarp Technology (Shanghai) Co., Ltd, Shanghai, China
Jesus Rico-Feijoo,
Hospital Universitario Río Hortega, Spain
Jian-cang Zhou,
Sir Run Run Shaw Hospital, China

****Correspondence:***
Huiying Zhao
zhaohuiying109@sina.com

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *14 September 2021*
Accepted: *18 October 2021*
Published: *05 November 2021*

Citation:
Zhao H, Kennedy JN, Wang S, Brant EB, Bernard GR, DeMerle K, Chang C-CH, Angus DC and Seymour CW (2021) Revising Host Phenotypes of Sepsis Using Microbiology. Front. Med. 8:775511. doi: 10.3389/fmed.2021.775511

Conclusions: Sepsis host phenotype assignment was significantly modified when microbiology data were added to clinical variables, increasing cluster cohesiveness and homogeneity.

Keywords: phenotype, latent class analysis, host, pathogen, sepsis

INTRODUCTION

There are more than 49 million worldwide cases of sepsis annually (1). Despite prompt recognition and treatment, sepsis remains a leading cause of mortality (2, 3). Many trials of candidate sepsis treatments failed to find beneficial effects, in part due to the wide heterogeneity in causative pathogens, host response, and patterns of organ dysfunction. A more precise treatment strategy is needed to move beyond a "one-size-fits-all" bundle (4–7).

Recent work proposed clinical and biologic phenotypes of sepsis that may identify groups for targeted treatment and enrichment strategies in clinical trials (8–17). These studies focus mainly on clinical data in the electronic health record (EHR), protein biomarkers, or molecular data. They do not typically incorporate microbiology or pathogen data as these features are (i) difficult to measure and adjudicate, and (ii) not available at the point-of-care. Despite the inclusion of causative pathogen in leading conceptual models of sepsis (18), its role in sepsis classification using machine learning is unknown.

To address this challenge, we performed a secondary analysis of the Recombinant Human Activated Protein C (rhAPC) Worldwide Evaluation in Severe Sepsis (PROWESS) Study, a large multicenter randomized clinical trial of sepsis patients unique for its detailed microbiology data (19). We aim to determine the effect of adding microbiology data to clinical sepsis phenotypes.

METHODS

The project was approved by the University of Pittsburgh institutional review board and conducted under data use agreements (PRO15110441 and PRO17120315). The original study was approved by the institutional review board at each site, and written informed consent was obtained. The informed consent specified that the data collected will be used for further scientific studies in addition to the original clinical trial (19).

Data and Study Population

We conducted a secondary analysis of the PROWESS study, which enrolled 1,690 patients with severe sepsis at 164 centers in 11 countries from July 1998 to June 2000. Severe sepsis was defined as a known or suspected infection, 3 or more signs of systemic inflammation, and the sepsis-induced dysfunction of at least one organ or system. Patients were enrolled within 24 h after they met the criteria of severe sepsis. Patients were randomly assigned 1:1 to receive drotrecogin alfa or placebo at each center within 24 h of meeting inclusion criteria (19).

Clinical and Microbiology Variables for Phenotyping

We selected 24 clinical variables prior to randomization and 3 microbiological variables for analysis. We used clinical variables previously mapped to sepsis phenotypes (15). They included demographic variables (e.g., age, sex, Elixhauser comorbidities), vital signs [e.g., heart rate, respiratory rate, Glasgow coma scale (GCS) score, systolic blood pressure (SBP), temperature, and oxygen saturation (SaO_2)], markers of inflammation [e.g., white blood cell count (WBC), premature neutrophil count ("bands")], markers of organ dysfunction or injury [e.g., alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, blood urea nitrogen (BUN), creatinine, partial pressure of oxygen (PaO_2), platelets, and prothrombin time]and serum glucose, sodium, hemoglobin, chloride, and albumin.

The microbiological variables in PROWESS included the site of infection (e.g., bloodstream, central nervous system, genitourinary, abdominal, lung, and others), type of pathogen identified from a positive culture (e.g., mixed, fungus, gram negative, gram positive, and organism negative), and drug resistance (one or more drug resistance vs. no drug resistance).

Biomarkers, Clinical Outcomes, and Treatment Effects

After phenotypes were assigned, we studied 14 serum biomarkers measured at baseline prior to randomization. They included inflammatory biomarkers [e.g., interleukin (IL)-1b, IL-6, IL-8, IL-10, and tumor necrosis factor (TNF)] and coagulation biomarkers [e.g., antithrombin, d-dimer, factor V, plasminogen activator inhibitor (PAI)-1, plasminogen activity, protein C activity, protein S activity, prothrombin fragment 1–2, and thrombin-antithrombin (TAT) complex].

The primary outcome was 28 day mortality. Secondary outcomes were 90 day mortality and 180 day mortality.

Statistical Methods

To derive phenotypes, we first explored candidate variable distributions, missingness (**Supplementary Table 1**), and correlation. We applied log transformations to non-normal data. We handled missing data by using multiple imputations by chained equations (MICE) (20). We included all covariates in the imputation procedure, and modeled variables using logistic, linear, multinomial, or ordinal regression, as appropriate. We evaluated distributions of clustering variables before and after imputation (**Supplementary Table 2**), and correlation of variables using rank order statistics (**Supplementary Figure 1**).

We used latent class analysis (LCA) to derive host (24 clinical variables) and host-pathogen (24 clinical plus 3 microbiological variables) phenotypes (21). We determined the optimal number

TABLE 1 | Characteristics of the host model phenotypes.

Characteristic	All patients	α-type	β-type	γ-type	δ-type	*P*-value*
No. of patients (%)	1,690	327 (19.4%)	518 (30.7%)	532 (31.5%)	313 (18.5%)	
Age, median [IQR], years	64 [49–74]	44 [34–55]	71 [63–77]	65 [52–74]	59 [46–73]	<0.01
Gender, no. (%)						0.25
Male	964 (57%)	174 (53%)	310 (60%)	307 (58%)	173 (55%)	
Female	726 (43%)	153 (47%)	208 (40%)	225 (42%)	140 (45%)	
Elixhauser Comorbidities, median [IQR]	1 [0–2]	0 [0–1]	2 [1–3]	1 [0–2]	1 [0–2]	<0.01
Inflammation						
Premature neutrophil count (bands), median [IQR], %	1.1 [0.5–2.6]	0.8 [0.3–1.4]	1.1 [0.6–2.2]	1.3 [0.5–3.5]	1.2 [0.4–2.7]	<0.01
Temperature, median [IQR], °C	38.6 [37.6–39.3]	39.0 [38.5–39.5]	38.1 [35.9–38.9]	38.7 [37.7–39.4]	38.6 [37.0–39.6]	<0.01
White blood cell count, median [IQR], $\times 10^9$/L	14 [9–20]	14 [9–18]	15 [10–21]	13 [7–21]	15 [8–22]	<0.01
Pulmonary						
Oxygen saturation, median [IQR], %	95 [90–97]	94 [88–97]	96 [92–98]	94 [90–96]	95 [89–98]	<0.01
Partial pressure of oxygen, arterial, median [IQR], mmHg	76 [62–101]	71 [55–92]	84 [65–125]	71 [62–82]	90 [63–132]	<0.01
Respiratory rate, median [IQR], breaths/min	31 [23–40]	32 [24–40]	28 [19–35]	32 [24–40]	32 [24–40]	<0.01
Cardiovascular or Hemodynamic						
Heart rate, median [IQR], beats/min	130 [115–147]	133 [120–148]	122 [105–140]	136 [123–150]	133 [115–150]	<0.01
Systolic blood pressure, median [IQR], mmHg	80 [70–95]	90 [80–110]	85 [69–103]	78 [68–86]	77 [65–92]	<0.01
Renal						
Blood urea nitrogen, median [IQR], mg/dL	10 [6–15]	5 [4–7]	11 [7–16]	11 [8–16]	14 [10–20]	<0.01
Creatinine, median [IQR], mg/dL	1.5 [1.0–2.3]	0.9 [0.7–1.1]	1.4 [1.0–2.1]	1.8 [1.3–2.5]	2.3 [1.6–3.4]	<0.01
Hepatic						
Alanine transaminase, median [IQR], U/L	28 [16–55]	26 [15–43]	20 [13–31]	27.5 [17–50]	130 [50–395]	<0.01
Aspartate transaminase, median [IQR], U/L	43 [24–93]	37 [22–68]	28 [20–43]	47 [28–85]	246 [102–616]	<0.01
Bilirubin, median [IQR], mg/dL	0.7 [0.4–1.3]	0.7 [0.4–1.3]	0.5 [0.3–0.9]	0.8 [0.5–1.5]	1.0 [0.6–2.2]	<0.01
Hematologic						
Hemoglobin, median [IQR], g/dL	11 [9–12]	11 [10–12]	10 [9–12]	11 [9–12]	11 [10–12]	0.02
Platelets, median [IQR], $\times 10^9$/L	168 [105–240]	193 [140–256]	205 [147–290]	135 [90–199]	129 [71–200]	<0.01
Prothrombin time, median [IQR], secs	19 [17–22]	17 [16–19]	18 [16–20]	20 [18–24]	22 [18–30]	<0.01
Other						
Albumin, median [IQR], g/dL	2.0 [1.6–2.4]	2.2 [1.7–2.6]	2.0 [1.6–2.5]	1.9 [1.5−2.3]	2.0 [1.5–2.5]	<0.01
Chloride, median [IQR], mEq/L	106 [101–111]	105 [102–110]	106 [100–111]	107 [103–112]	105 [100–111]	<0.01
Glasgow Coma Scale score, median [IQR]	14 [11–15]	15 [12–15]	14 [9–15]	15 [12–15]	14 [10–15]	<0.01
Glucose, median [IQR], mg/dL	146 [115–196]	133 [112–162]	163 [124–227]	144 [112–198.5]	144 [108–196]	<0.01
Sodium, median [IQR], mEp/L	139 [135–143]	139 [135–142]	139 [135–143]	139 [136–142]	139 [135–144]	0.34
Outcomes						
28 day mortality, no. (%)	469 (28%)	29 (9%)	157 (30%)	152 (29%)	131 (42%)	<0.01
90 day mortality, no. (%)	593 (35%)	43 (13%)	210 (41%)	188 (35%)	152 (49%)	<0.01
180 day mortality, no. (%)	638 (38%)	51 (16%)	233 (45%)	198 (37%)	156 (50%)	<0.01

**Kruskal-Wallis used for continuous and or chi-square for categorical comparisons, across four phenotypes. IQR, interquartile range.*

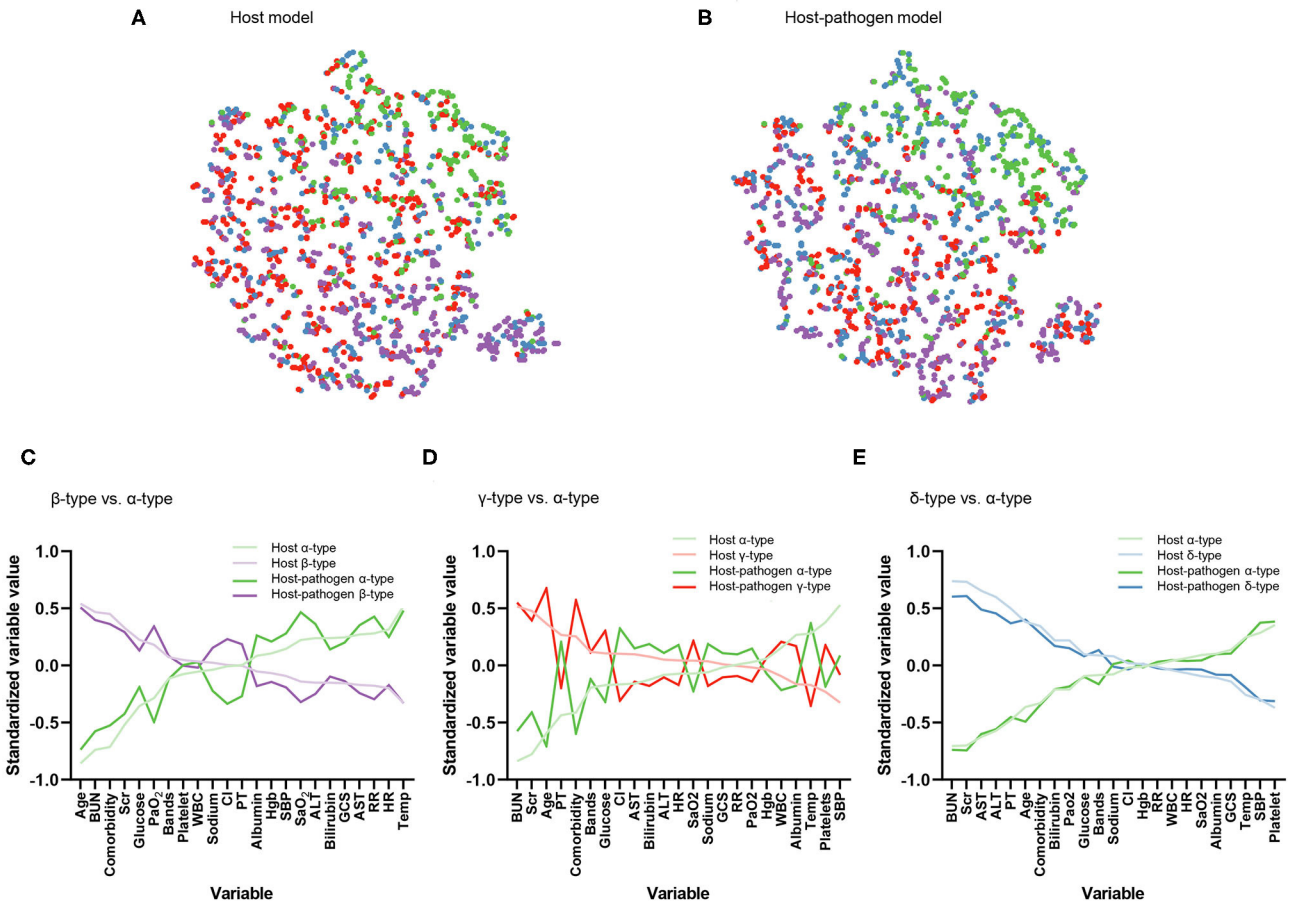


FIGURE 1 | Visualization of phenotype assignments and comparison of clinical variables that contribute to phenotypes. **(A)** Visualization of phenotype assignments in host, **(B)** host-pathogen model using t-distributed stochastic neighbor embedding (t-SNE) plot. Green, purple, red, and blue dots represent α-type, β-type, γ-type and δ-type, respectively. Phenotype members have a similar frequency and distribution across models. **(C)** Differences in standardized mean value of each variable ranked from maximum positive to negative separation (x-axis). Dark lines correspond to host model. Light lines correspond to same comparisons but from host-pathogen model. Plot compares β-type (purple) to α-type (green). Variables ranked on the left x-axis are greater in β-type than α-type (e.g., age, BUN, and comorbidity) while those on the right x-axis are lower in β-type than α-type (e.g., temperature, heart rate). **(D)** Comparison between γ-type (red) and α-type (green). **(E)** Comparison between δ-type (blue) and α-type (green).

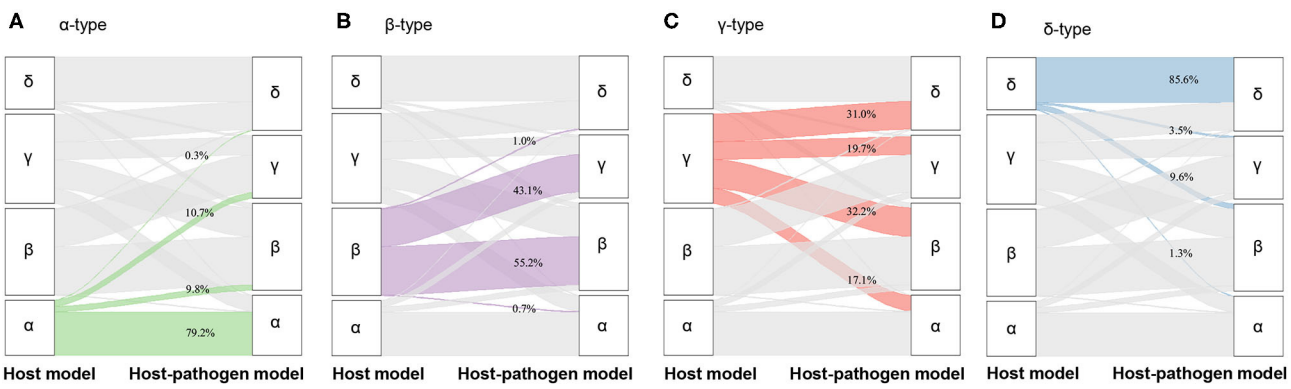


FIGURE 2 | Alluvial plot showing the change of membership from host model to host-pathogen model. **(A)** The change of membership from α-type of host model (green, left column, *N* = 327) to host-pathogen model (right column), **(B)** from β-type of host model (purple, left column, *N* = 518) to host-pathogen model (right column), **(C)** from γ-type of host model (red, left column, *N* = 532) to host-pathogen model (right column), **(D)** from δ-type of host model (blue, left column, *N* = 313) to host-pathogen model (right column).

of phenotypes using the minimum Bayesian information criteria (BIC), class size, median probabilities of group membership, entropy, and clinical features of groups. For each patient, we used LCA to produce a posterior probability describing the likelihood of the patient belonging to the phenotype, with posterior probability ranges from 0 to 1. Patients were assigned to the phenotype for which they had the highest posterior probability. We estimated models ranging from two to seven phenotypes (**Supplementary Table 3**). We determined the optimal number of clusters using a combination of criteria, (i) a smaller BIC, (ii) a higher Entropy, (iii) adequate sample size within cluster, (iv) higher median posterior probabilities of group membership, and (v) clinical characteristics of the clusters. We illustrated the host and host-pathogen models in 2 ways: (i) t-distributed stochastic neighbor embedding (t-SNE) plots (which show the frequency and distribution of phenotype members) and (ii) alluvial plots (which show the change of membership between host and host-pathogen models by phenotypes). We compared the contribution of continuous variables to phenotypes in both host and host-pathogen models using the differences in standardized mean value of each variable.

To quantify the change in phenotypes after addition of microbiology, we measured the mean (SD) probabilities of membership for the assigned group(s). We also compared the proportion of patients in each group using chi square tests. We tested for differences in 28, 90 and 180 day mortality between phenotypes using chi square and Kaplan-Meier curves to illustrate differences in 28 day mortality. We tested the treatment effects for rhAPC by phenotype using Kaplan-Meier curves of 28 day mortality. We conducted 2 sensitivity analyses, (i) excluding variables with high missingness (missing >50%: hemoglobin and premature neutrophil count [bands]) and (ii) using a 5-class model as the optimal fit for the LCA. Analyses were performed with Stata 15.1 (StataCorp, College Station, Texas), and R 3.4.1 (depmixS4 package for LCA; Rtsne package for making t-SNE plots; alluvial package for making alluvial plots, Version: 0.1-2. Bojanowski M and Edwards R; 2016. https://github.com/mbojan/alluvial) with a significance threshold of <0.05 in 2-sided tests.

RESULTS

Patients

Among 1,690 subjects, the median age was 64 [IQR: 49–74] years old, 964 (57%) patients were male, and median Elixhauser comorbidity index was 1 [IQR: 0–2] (**Table 1**, **Supplementary Table 4**). The primary infection site was lung (54%), compared to abdominal (19%) or genitourinary (11%) infections. A mixed pathogen infection (35%) was the most common, compared to gram positive (22%) or gram negative bacteria alone (16%).

Host Model

Using 24 clinical variables in the latent class analysis (host model), we determined that a 4-class model was the optimal fit [applied labels alpha (α), beta (β), gamma (γ), and delta (δ) types]. Entropy in all models was 0.75 or greater, and the BIC decreased

TABLE 2 | Example characteristics of β-type and γ-type in host and host-pathogen models.

Variable	Host β-type	Host-pathogen β-type	Host γ-type	Host-pathogen γ-type
No. of patients (%)	518 (31%)	519 (31%)	532 (32%)	374 (22%)
Clinical variable				
Age, median [IQR]	71 [63–77]	69 [58–77]	65 [52–74]	70 [61–77]
Elixhauser comorbidity, median [IQR]	2 [1–3]	1 [1–2]	1 [0–2]	2 [1–3]
Heart rate, beats/min, median [IQR]	122 [105–140]	126 [110–144]	136 [123–150]	127 [112–145]
SBP, mmHg, median [IQR]	85 [69–103]	80 [70–92]	78 [68–86]	83 [69–103]
Bilirubin, mg/dL, median [IQR]	0.5 [0.3–0.9]	0.6 [0.4–1.1]	0.8 [0.5–1.5]	0.6 [0.3–0.9]
Glucose, mg/dL, median [IQR]	163 [124–227]	147 [117–198]	144 [112–199]	166 [127–227]
Oxygen saturation, %, median [IQR]	96 [92–98]	97 [94–98]	94 [90–96]	92 [85–95]
PaO_2, mmHg, median [IQR]	84 [65–125]	92 [73–138]	71 [62–82]	64 [53–77]
Platelets, $\times 10^9$/L, median [IQR]	205 [147–290]	175 [116–252]	135 [90–199]	205 [157–281]
Prothrombin time, s, median [IQR]	18 [16–20]	19 [17–23]	20 [18–24]	17 [15–19]
WBC Count, $\times 10^9$/L, median [IQR]	15 [10–21]	13 [8–19]	13 [7–21]	16 [12–21]
Microbiological variable				
Source				
Bloodstream, no. (%)	17 (3.3%)	16 (3.1%)	27 (5.1%)	2 (0.5%)
Abdominal, no. (%)	102 (20%)	208 (40%)	120 (23%)	8 (2.1%)
Lung, no. (%)	293 (57%)	183 (35%)	268 (50%)	291 (78%)
Type				
Mixed, no. (%)	188 (36%)	230 (44%)	183 (34%)	96 (26%)
Gram positive, no. (%)	87 (17%)	87 (17%)	130 (24%)	67 (18%)
Organism negative, no. (%)	128 (25%)	90 (17%)	109 (21%)	129 (35%)
Drug resistance, no. (%)	133 (32%)	198 (38%)	116 (28%)	64 (17%)

IQR, interquartile range; PaO_2, partial pressure of oxygen; SBP, systolic blood pressure; WBC, white blood cell.

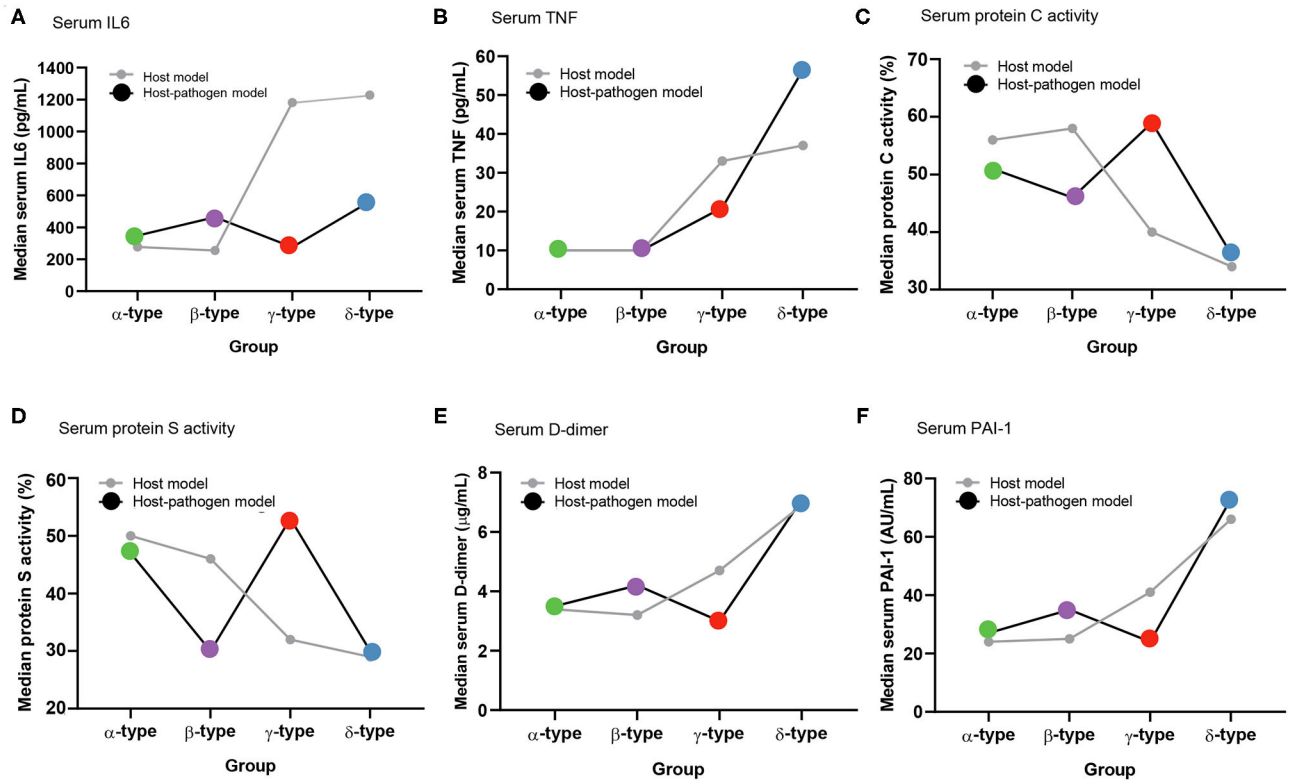


FIGURE 3 | Comparison of median values of biomarkers between host and host-pathogen models across phenotypes. **(A)** Serum IL-6, **(B)** serum TNF-α, **(C)** serum protein C activity, **(D)** serum protein S activity, **(E)** serum D-dimer, and **(F)** serum plasminogen activator inhibitor-1 (PAI-1). Gray lines represent host model, and black lines represent host-pathogen model. Green, purple, red, and blue dots represent α-type, β-type, γ-type and δ-type, respectively.

as class number increased from 2 to 4. The median probability of group membership was high (>95%, **Supplementary Table 3**, **Supplementary Figure 2**). Phenotypes ranged in size from 19 to 32% of the cohort, and differed broadly in clinical characteristics (**Table 1**, **Figure 1**). Consistent with prior data (15), patients with the α-type (19%) were younger and had less shock, β-type (31%) were older and had greater comorbidity, γ-type (32%) had more pulmonary dysfunction, and δ-type (19%) had more liver, renal, and hematologic dysfunction and shock.

Host Pathogen Model

When 3 microbiological variables were included in the latent class analysis (host-pathogen model), a 4-class model again demonstrated optimal fit (also applied labels α, β, γ, and δ types) (**Supplementary Table 3**, **Supplementary Figure 2**). We visualized patients using t-SNE plots (**Figures 1A,B**) and found that the proportion of phenotype members was similar in host and host-pathogen models. However, 772 of 1,690 (46%) patients changed phenotypes, particularly the β (45%) and γ-types (80%) (**Figure 2**, **Supplementary Table 5**). The host-pathogen phenotypes had higher median membership probabilities than host phenotypes alone (host: 0.95 vs. host-pathogen: 0.97, $P < 0.01$, **Supplementary Table 5**). Among patients who rearranged phenotypes in the host-pathogen model, the initial host model membership probability was lower than patients who did not change (median 0.90 vs. 0.98, $p < 0.01$, **Supplementary Table 6**).

The contribution of individual variables to phenotypes are ranked before and after including microbiology data. These plots show little change for δ- and α-types, but greater inconsistency for the β- and γ-type variables (**Figures 1D–F**). For example, among β-type patients, the proportion of abdominal infections (from 20 to 40%) and mixed-type infections (from 36 to 44%) increased, while the proportion of lung infections decreased from 57 to 35%; γ-type patients had an increased rate of lung infections (from 50 to 78%) with worsening pulmonary function (PaO_2 decreased from 71 to 64 mmHg) (**Tables 1**, **2**, **Supplementary Tables 4**, **7**, **8**).

Correlation With Baseline Biomarkers and 28-Day Mortality

Comparing host and host-pathogen models, 13 of 14 biomarkers were significantly different across phenotypes when adding microbiology data (excluding only IL-1b, **Supplementary Tables 9, 10**). For example, in the β-type, the median level of PAI-1 increased from 25 to 35 AU/mL, and d-dimer increased from 3.2 to 4.2 μg/mL; while PAI-1 (from 41 to 24 AU/mL) and d-dimer (from 4.7 to 3.0 μg/mL) decreased in the γ-type (**Figure 3**). The cumulative 28 day mortality

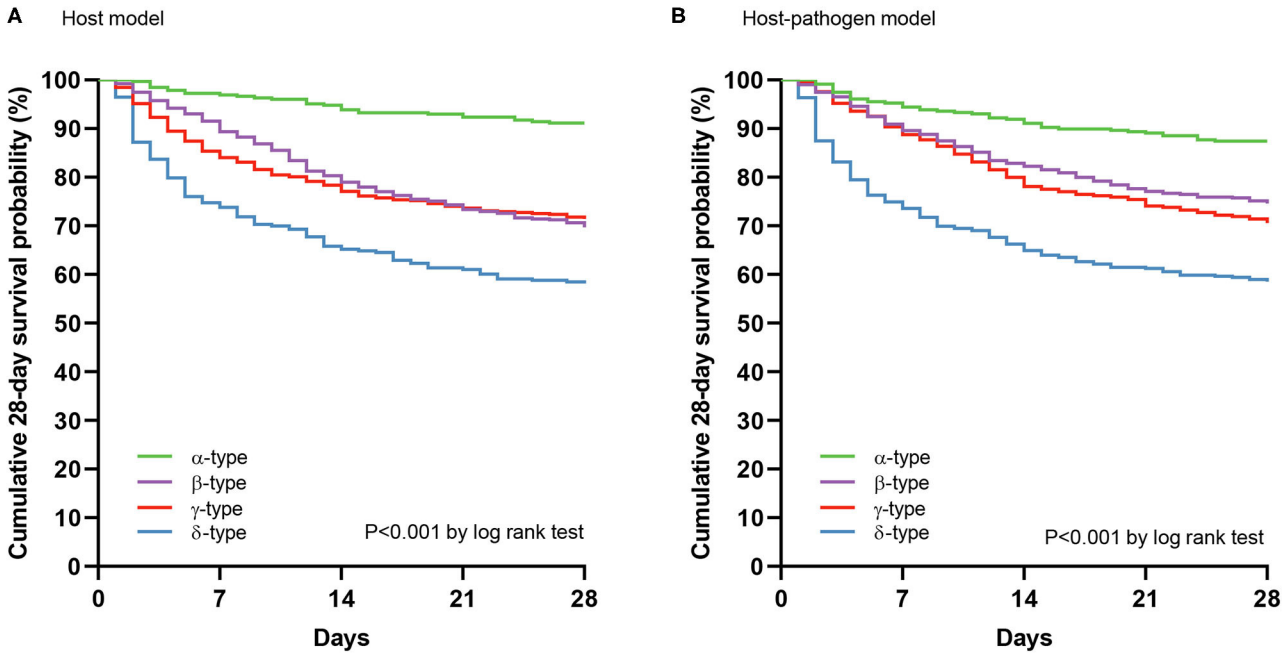


FIGURE 4 | 28 day mortality by phenotypes using Kaplan-Meier curves. **(A)** Cumulative survival at 28 days by phenotypes using Kaplan-Meier curves in host model, and **(B)** host-pathogen model. Green, purple, red, and blue lines represent α-type, β-type, γ-type and δ-type, respectively.

probability was significantly different for individual phenotypes in host and host-pathogen models (both log-rank $P < 0.01$), but was similar between models. In both models, 90 day and 180 day mortality were also significantly different for individual phenotypes (all chi-square $P < 0.01$), but were similar between models (**Figure 4**; **Table 1**, **Supplemental Tables 4**, **7**).

Treatment Effect for rhAPC by Phenotype After Including Microbiology Variables

In host model, rhAPC significantly decreased the cumulative 28 day mortality probability in gamma type ($P = 0.047$ by log rank test), while when microbiology variables were added, the 28 day mortality was similar between rhAPC and placebo group ($P = 0.780$ by log rank test) (**Figure 5**).

Sensitivity Analysis

To understand the robustness of these results, we derived phenotypes excluding variables with high missingness and found that a 4-class model remained optimal for both host and host-pathogen models (**Supplementary Figure 3**). In addition, these models had similar frequency and characteristics to phenotypes as the primary analysis (**Supplementary Table 11**, **Supplementary Figure 3**). For example, 713 (42%) patients were rearranged when microbiological variables were added, with highest rates of change in the β and γ-type (**Supplementary Figure 4**). We also explored a 5-class model and found that microbiological variables also rearranged 632 (37%) of patients, increased the probability of membership, and changed variable characteristics in clinically meaningful way (**Supplementary Tables 12**, **13**, **Supplementary Figure 5**).

DISCUSSION

In this proof-of-concept analysis, the addition of microbiological variables to host sepsis phenotypes led to meaningful rearrangement of patients, particularly the beta and gamma types. These changes did not modify short or long-term outcomes, but changed the treatment effect for rhAPC in gamma type. This work suggests that pathogen data may have an under-recognized role in sepsis phenotype classification using machine learning methods.

For decades, sepsis has been characterized by the offending pathogen, such as Neisseria meningitis or pneumococcal pneumonia. However, these labels alone do not capture the combined complexity of the host response, tolerance, or damage in sepsis (22). Recent work using machine learning to subtype sepsis did not include pathogen data due to practical measurement challenges during emergency care (8–11, 15, 23). Preliminary work in the PROWESS-SHOCK trial began to use microbiology together with clinical data to propose subphenotypes of septic shock (17). We extend this work by investigating the question, how much does microbiology add beyond that of clinical data alone? This is a key knowledge gap that will guide the embedding of sepsis phenotypes into trials and clinical practice.

We found that the addition of microbiological variables to host phenotypes led to meaningful rearrangement of sepsis patients. A large proportion, particularly of the gamma type, were assigned to a different phenotype. The host pathogen model also appeared to statistically increase in probability of assignment. These changes were not, however, accompanied by changes

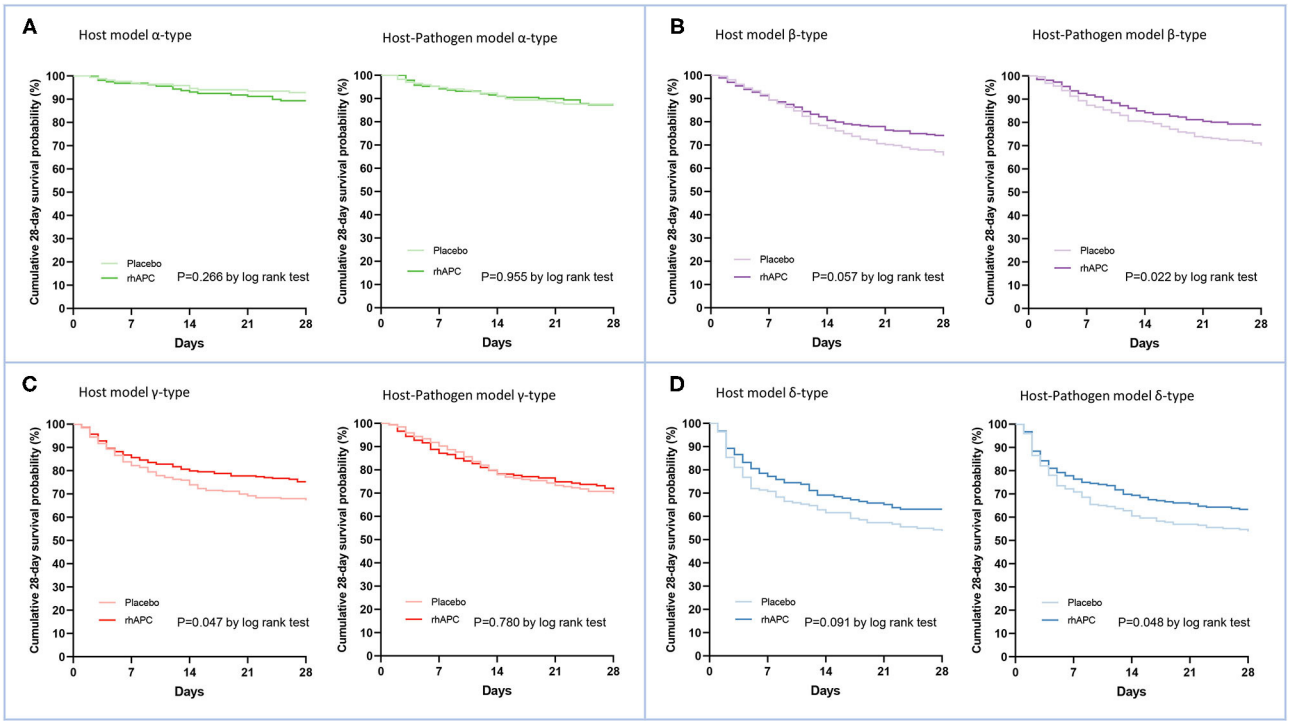


FIGURE 5 | Comparison of the treatments effects for Recombinant Human Activated Protein C by phenotype. **(A)** The comparison of cumulative survival at 28 days using Kaplan-Meier curves between rhAPC group and placebo group in α-type of both host and host-pathogen model, **(B)** β-type, **(C)** γ-type, and **(D)** δ-type. Dark lines correspond to rhAPC group and light lines correspond to placebo group. Green, purple, red, and blue dots represent α-type, β-type, γ-type and δ-type, respectively.

in patient outcomes by phenotype. We also found that the addition of pathogen data could obviously change the treatment effect for rhAPC in gamma phenotype. It further elaborated the importance of pathogen data to sepsis phenotyping. As a proof of concept analysis, many important steps follow, (i) to reproduce in larger, generalizable cohort; (ii) determine if other treatment effects, perhaps time to antimicrobials or source control, are modified by pathogen informed subtypes.

A challenge to the incorporation of microbiological data into sepsis phenotypes is that these parameters are not routinely available during emergency care or at the time of typical enrollment in clinical trials. Several rapid approaches are under study to identify infection type (e.g., bacterial, viral), or drug resistance. These include multiplex real-time polymerase chain reaction (PCR) systems, next-generation sequencing (NGS) (24–26), and those probing the pathogen specific host response (27, 28). These approaches have complex workflow, a need for rigorous quality control, and a yet-to-be-determined optimal "clinical moment" in bedside care.

This study has several limitations. First, we performed a proof of concept in a single trial with small sample, and generalizability requires further study. Second, the microbiology data were derived from the culture results of the database of PROWESS which could not accurately and completely distinguish the colonization, positive cultured infection, and negative cultured infection. In addition, due to the low incidence, we did not identify multidrug-resistant (MDR) and extensively drug-resistant (XDR) bacteria in the drug resistance variables, these two variables have greater clinical application value. Third, most pathogens were bacteria, with low rates of viral and fungal infection. Additional data is needed to parse through the role of specific viral pathogens to phenotypes. Fourth, missing data were common. Although we used multiple imputation, bias may be introduced for those variables with high missingness. To address this limitation, we excluded variables with high missingness (>50%) in sensitivity analyses and found similar results. Fifth, we compared mortality and treatment effects of rhAPC between host and host-pathogen models using Kaplan-Meier curves which may lead to non-rigorous results. Further need to verify these effects using stratified proportional hazards model in larger sample study. Sixth, the choice of optimal number of clusters is semi-subjective and different statistical approaches are available to determine cluster number. Informed by prior work in SENECA (15), we focused on 4 class models. However, we explored a 5-class model in sensitivity analyses and found similar trends to those observed in the primary analysis.

CONCLUSION

Sepsis host phenotype assignment was significantly modified when microbiology data were added to clinical variables, increasing cluster cohesiveness and homogeneity. The clinical

significance of these changes and importance for treatment effects in clinical trials remains uncertain.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of Pittsburgh institutional review board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

HZ, JK, DA, and CS contributed to study conception and design. GB, DA, and CS contributed to acquisition of data. HZ, JK, SW, EB, KD, C-CC, DA, and CS contributed to analysis and interpretation of data. HZ and CS drafted the manuscript. CS supervised the study. All authors critically revised the manuscript.

FUNDING

HZ was supported in part by grant from Peking University People's Hospital Research and Development Funds (RDY2019-43, derive sepsis phenotypes using electronic medical data and machine learning). CS was supported in part by grants from the National Institutes Health (R35GM119519).

ACKNOWLEDGMENTS

We acknowledge the contribution of the patients, families, researchers, and clinical staff for the data included in this study. We acknowledge Eli Lilly Inc for providing the PROWESS trial dataset.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.775511/full#supplementary-material

REFERENCES

1. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *Lancet.* (2020) 395:200–11. doi: 10.1016/S0140-6736(19)32989-7
2. Stevenson EK, Rubenstein AR, Radin GT, Wiener RS, Walkey AJ. Two decades of mortality trends among patients with severe sepsis: a comparative meta-analysis. *Crit Care Med.* (2014) 42:625–31. doi: 10.1097/CCM.0000000000000026
3. Levy MM, Rhodes A, Phillips GS, Townsend SR, Schorr CA, Beale R. et al. Surviving sepsis campaign: association between performance metrics and outcomes in a 75-year study. *Intensive Care Med.* (2014) 40:1623–33. doi: 10.1007/s00134-014-3496-0
4. Angus DC. The search for effective therapy for sepsis: back to the drawing board? *JAMA.* (2011) 306:2614–5. doi: 10.1001/jama.2011.1853
5. Angus DC, van der Poll T. Severe sepsis and septic shock. *N Engl J Med.* (2013) 369:840–51. doi: 10.1056/NEJMra1208623
6. Seymour CW, Gomez H, Chang CH, Clermont G, Kellum JA, Kennedy J, et al. Precision medicine for all? challenges and opportunities for a precision medicine approach to critical illness. *Crit Care.* (2017) 21:257. doi: 10.1186/s13054-017-1836-5
7. Vincent JL. The coming era of precision medicine for intensive care. *Crit Care.* (2017) 21:314. doi: 10.1186/s13054-017-1910-z
8. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med.* (2014) 2:611–20. doi: 10.1016/S2213-2600(14)70097-9
9. Calfee CS, Delucchi KL, Sinha P, Matthay MA, Hackett J, Shankar-Hari M, et al. Acute respiratory distress syndrome subphenotypes and differential response to simvastatin: secondary analysis of a randomised controlled trial. *Lancet Respir Med.* (2018) 6:691–8.
10. Famous KR, Delucchi K, Ware LB, Kangelaris KN, Liu KD, Thompson BT, et al. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med.* (2017) 195:331–8. doi: 10.1164/rccm.201603-0645OC
11. Sinha P, Delucchi KL, Thompson BT, McAuley DF, Matthay MA, Calfee CS, et al. Latent class analysis of ARDS subphenotypes: a secondary analysis of the statins for acutely injured lungs from sepsis (SAILS) study. *Intensive Care Med.* (2018) 44:1859–69. doi: 10.1007/s00134-018-5378-3
12. Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, et al. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. *Lancet Respir Med.* (2016) 4:259–71. doi: 10.1016/S2213-2600(16)00046-1
13. Sweeney TE, Azad TD, Donato M, Haynes WA, Perumal TM, Henao R, et al. Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Crit Care Med.* (2018) 46:915–25. doi: 10.1097/CCM.0000000000003084
14. Zhang Z, Zhang G, Goyal H, Mo L, Hong Y. Identification of subclasses of sepsis that showed different clinical outcomes and responses to amount of fluid resuscitation: a latent profile analysis. *Crit Care.* (2018) 22:347. doi: 10.1186/s13054-018-2279-3
15. Seymour CW, Kennedy JN, Wang S, Chang CH, Elliott CF, Xu Z, et al. Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA.* (2019) 321:2003–17. doi: 10.1001/jama.2019.5791
16. Knox DB, Lanspa MJ, Kuttler KG, Brewer SC, Brown SM. Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. *Intensive Care Med.* (2015) 41:814–22. doi: 10.1007/s00134-015-3764-7
17. Gardlund B, Dmitrieva NO, Pieper CF, Finfer S, Marshall JC, Thompson BT. Six subphenotypes in septic shock: latent class analysis of the PROWESS shock study. *J Crit Care.* (2018) 47:70–9. doi: 10.1016/j.jcrc.2018.06.012
18. Medzhitov R, Schneider DS, Soares MP. Disease tolerance as a defense strategy. *Science.* (2012) 335:936–41. doi: 10.1126/science.1214935
19. Bernard GR, Vincent JL, Laterre PF, LaRosa SP, Dhainaut JF, Lopez-Rodriguez A, et al. Efficacy and safety of recombinant human activated protein C for severe sepsis. *N Engl J Med.* (2001) 344:699–709. doi: 10.1056/NEJM200103083441001
20. Newgard CD, Haukoos JS. Advanced statistics: missing data in clinical research–part 2: multiple imputation. *Acad Emerg Med.* (2007) 14:669–78. doi: 10.1111/j.1553-2712.2007.tb01856.x

21. Rindskopf D, Rindskopf W. The value of latent class analysis in medical diagnosis. *Stat Med.* (1986) 5:21–7. doi: 10.1002/sim.4780050105
22. Goh C, Knight JC. Enhanced understanding of the host-pathogen interaction in sepsis: new opportunities for omic approaches. *Lancet Respir Med.* (2017) 5:212–23. doi: 10.1016/S2213-2600(17)30045-0
23. Santhakumaran S, Gordon A, Prevost AT, O'Kane C, McAuley DF, Shankar-Hari M. Heterogeneity of treatment effect by baseline risk of mortality in critically ill patients: re-analysis of three recent sepsis and ARDS randomised controlled trials. *Crit Care.* (2019) 23:156. doi: 10.1186/s13054-019-2446-1
24. Dark P, Blackwood B, Gates S, McAuley D, Perkins GD, McMullan R, et al. Accuracy of LIGHTCYCLER((R)) septifast for the detection and identification of pathogens in the blood of patients with suspected sepsis: a systematic review and meta-analysis. *Intensive Care Med.* (2015) 41:21–33. doi: 10.1007/s00134-014-3553-8
25. Maugeri G, Lychko I, Sobral R, Roque ACA. Identification and antibiotic-susceptibility profiling of infectious bacterial agents: a review of current and future trends. *Biotechnol J.* (2019) 14:e1700750. doi: 10.1002/biot.201700750
26. Gwinn M, MacCannell D, Armstrong GL. Next-generation sequencing of infectious pathogens. *JAMA.* (2019) 321:893–4. doi: 10.1001/jama.2018.21669
27. Sweeney TE, Liesenfeld O, May L. Diagnosis of bacterial sepsis: why are tests for bacteremia not sufficient? *Expert Rev Mol Diagn.* (2019) 19:959–62. doi: 10.1080/14737159.2019.1660644
28. Sweeney TE, Wong HR, Khatri P. Robust classification of bacterial and viral infections via integrated host gene expression diagnostics. *Sci Transl Med.* (2016) 8:346ra91. doi: 10.1126/scitranslmed.aaf7165

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 08 November 2021
doi: 10.3389/fmed.2021.785711

Rule-Based Models for Risk Estimation and Analysis of In-hospital Mortality in Emergency and Critical Care

Oliver Haas[1,2]*, *Andreas Maier*[2] *and Eva Rothgang*[1]

[1] *Department of Industrial Engineering and Health, Institute of Medical Engineering, Technical University Amberg-Weiden, Weiden, Germany,* [2] *Pattern Recognition Lab, Department of Computer Science, Technical Faculty, Friedrich-Alexander University, Erlangen, Germany*

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Cledy Eliana Santos,
Community Health Service of Hospital Conceição Group, Brazil
Rui Guo,
First Affiliated Hospital of Chongqing Medical University, China

***Correspondence:**
Oliver Haas
o.haas@oth-aw.de

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *29 September 2021*
Accepted: *18 October 2021*
Published: *08 November 2021*

Citation:
Haas O, Maier A and Rothgang E (2021) Rule-Based Models for Risk Estimation and Analysis of In-hospital Mortality in Emergency and Critical Care. Front. Med. 8:785711. doi: 10.3389/fmed.2021.785711

We propose a novel method that uses associative classification and odds ratios to predict in-hospital mortality in emergency and critical care. Manual mortality risk scores have previously been used to assess the care needed for each patient and their need for palliative measures. Automated approaches allow providers to get a quick and objective estimation based on electronic health records. We use association rule mining to find relevant patterns in the dataset. The odds ratio is used instead of classical association rule mining metrics as a quality measure to analyze association instead of frequency. The resulting measures are used to estimate the in-hospital mortality risk. We compare two prediction models: one minimal model with socio-demographic factors that are available at the time of admission and can be provided by the patients themselves, namely gender, ethnicity, type of insurance, language, and marital status, and a full model that additionally includes clinical information like diagnoses, medication, and procedures. The method was tested and validated on MIMIC-IV, a publicly available clinical dataset. The minimal prediction model achieved an area under the receiver operating characteristic curve value of 0.69, while the full prediction model achieved a value of 0.98. The models serve different purposes. The minimal model can be used as a first risk assessment based on patient-reported information. The full model expands on this and provides an updated risk assessment each time a new variable occurs in the clinical case. In addition, the rules in the models allow us to analyze the dataset based on data-backed rules. We provide several examples of interesting rules, including rules that hint at errors in the underlying data, rules that correspond to existing epidemiological research, and rules that were previously unknown and can serve as starting points for future studies.

Keywords: in-hospital mortality, critical care, odds ratio, associative classification, machine learning, artificial intelligence

1. INTRODUCTION

The term in-hospital mortality defines the death of a patient during their stay at the hospital. Especially in emergency and critical care, many patients die in the hospital. While not all of these deaths can be prevented, early knowledge of a patient's in-hospital mortality risk can be used to assess the patient's status and necessary adjustments to this patient's care, reducing missed care and decreasing mortality rates (1). Apart from individual changes like the start of palliative care, organizational changes like a different allocation of nurse time or other resources can be informed by such risk scores. In-hospital mortality rates have also been used in the assessment of hospital care quality, as is the case in the United Kingdom (2).

In-hospital mortality risks are commonly determined manually (3). Manual scoring systems are built upon expert knowledge and have gone through significant development time. Another approach is Machine Learning (ML), which uses data and statistical methods to build a predictive model (4). Several methods for ML-based methods have been developed in recent years (5–7). While these approaches offer data-based evidence that is independent of expert knowledge, they face two challenges.

First, they often lack interpretability. As critical care is a life-and-death situation, providers need to be able to understand why a patient's status is assessed the way it is. Interpretability in Machine Learning is a broad field. Two distinctions to be made are local vs. global interpretability, i.e., whether the interpretation concerns one observation or the whole population, and model-specific vs. model-agnostic interpretability, i.e., whether the interpretation comes from within a specific model or is built on top of an existing model (8). We consider global model-specific interpretability. This offers two decisive advantages. First, global interpretability allows us to not only make predictions based on data, but also to explore the complex and heterogeneous data underlying our prediction model. Second, model-specific interpretability allows us to explain the reasoning behind our model and its inner workings to providers. Interpretability in in-hospital mortality risk estimation has previously been discussed (6, 9) and a trade-off between a prediction model's interpretability and predictive performance has been identified. Among the commonly used algorithms, Decision Trees (10) often offer high predictive performance while being highly interpretable (6).

The second challenge arises from the high number of possible variables in in-hospital mortality risk estimation. As critical care is complex, many variables can potentially play a role in estimating a patient's in-hospital mortality risk. This renders many common ML algorithms challenging to use and increases models' complexity, further hindering interpretability. Manual methods use expert knowledge from years of scientific research to identify which variables to include. Variables that were previously not considered but are readily available could play a role in in-hospital mortality risk estimation because they are highly correlated.

Association rule mining (ARM) is often used to detect patterns in high-dimensional data (11). This data mining method analyzes a given dataset for rules of the form "$A \Rightarrow B$," where A and B are sets of *items*, which in our case describe variable-value pairs. Such a rule denotes that in an observation in the dataset in which the items in A occur, the items in B will also likely occur. This algorithm produces a set of rules that fulfill pre-configured quality constraints. In the neighboring field of associative classification (AC), this class of algorithms is used to mine rules that help in the classification task at hand (12). AC algorithms first mine rules in which the right-hand side is the outcome of interest and then build a classifier based on these rules, e.g., by using the best rule that applies to an observation or by aggregating all applying rules (12). This leads to predictive models that are easy to interpret and offer a human-readable, model-specific, and global interpretation. Additionally, the rules in the model can be used to analyze the dataset itself due to their statistical nature. This helps detect interesting patterns in the data that correspond to correlations between clinical variables and the outcome of interest.

AC methods have previously been used in various fields of healthcare, including the prediction of outcomes and adverse events (13–16), the prediction of diseases and wellness (17–20), as well as biochemistry and genetics (21–25). AC methods have also been used in the field of in-hospital mortality risk estimation using the results of 12 lab tests (26). This shows the feasibility of AC methods in in-hospital mortality risk estimation.

We aim to improve automated, ML-based in-hospital mortality risk estimation methods by including heterogeneous variables as well as more variables in general. ARM methods, and thus also AC models, can incorporate large numbers of variables of different types, which is why we analyze the feasibility of AC models for in-hospital mortality risk estimation. We propose to use AC models to estimate the risk for in-hospital mortality risk estimation in critical and intensive care. The goal of the present study is 1) to analyze whether this approach is feasible and 2) which kind of rules are found to by such methods and which variables play a role in the prediction. This expands our previous work (27) in major ways. First, we present two models to analyze the temporal evolution of the prediction. Second, we expand our analysis of the resulting rules. Third, we compare our models to Decision Tree models. Lastly, we add more experiments to gain more insight into the model size.

2. MATERIALS AND METHODS

The method presented in this section was developed in the programming language C#. The overall concept of the proposed method, as well as a comparison to Decision Trees, can be seen in **Figure 1**. Based on a publicly available clinical dataset, we first mine association rules from the data. The resulting rules are then combined into a prediction model that can be applied to previously unseen cases. Finally, we test and validate the proposed method using experiments.

2.1. Dataset

We used data from the MIMIC-IV project, version 0.4 (28). MIMIC-IV is a collection of around 525,000 emergency department and intensive care unit cases, collected between

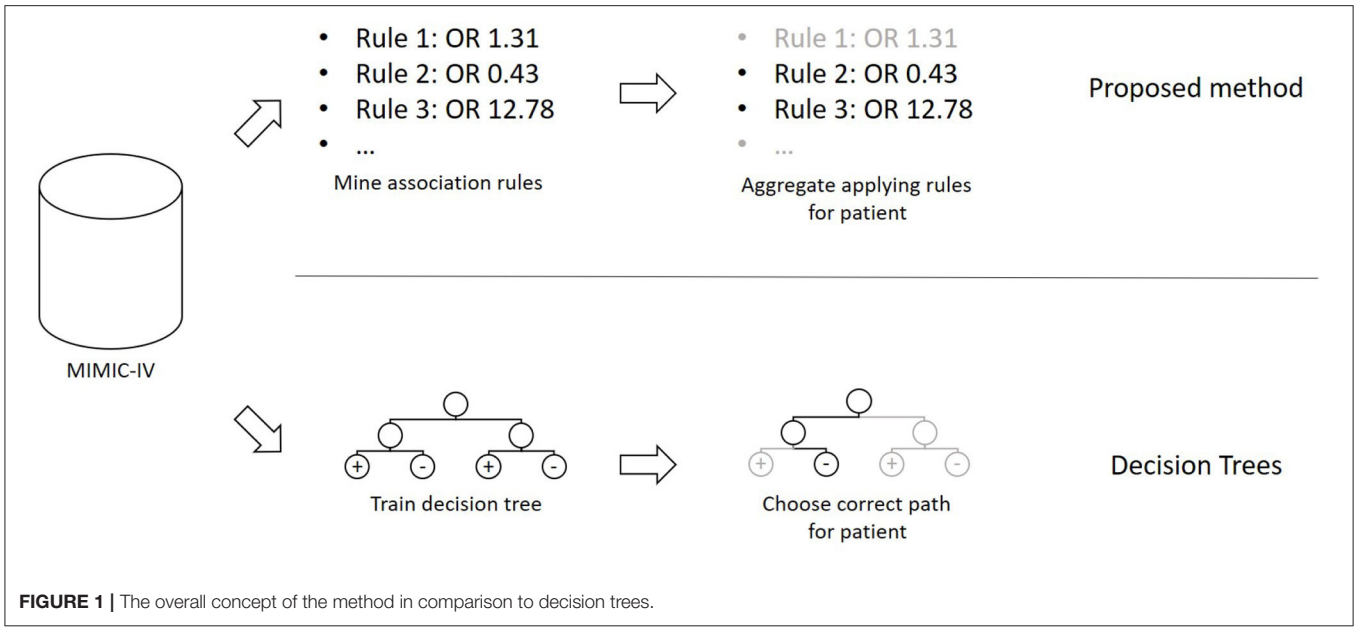


FIGURE 1 | The overall concept of the method in comparison to decision trees.

TABLE 1 | A list of all variables from MIMIC-IV that were used in this study.

Variable type	Minimal model?	Description	# variables
Diagnosis		Diagnoses (coded as ICD)	86,751
Ethnicity	✓	White, Asian, Black/African American etc.	8
Gender	✓	Binary: male/female	2
Insurance type	✓	Medicare, Medicaid, or Other	3
Language	✓	Binary: English/other	2
Marital status	✓	Single, married, divorced, widowed, or missing	5
Prescription		Drugs described to a patient	10,259
Procedure		Procedures (coded as ICD)	82,763
Service		Clinical services: neonatal, psychological etc.	21
Ward		Clinical wards: intensive care unit, surgery etc.	43
Overall	5 of 10		179,857

The table includes the name of the variable type, whether it is part of the minimal model, a description, as well as the number of variables. ICD International Classification of Diseases.

2008 and 2019 at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, United States of America. Recorded variables include diagnoses, procedures, drug prescriptions, socio-demographic factors like gender, insurance type, and marital status, and organizational information like diagnosis-related groups (DRGs), wards, and services.

Table 1 lists all types of variables used in this study. We included all categorical information that can be assumed to be available in most clinical contexts. This excludes rapidly changing information like vital parameters as well as textual notes. Not all of this information is available at the beginning of a clinical case. Some information, like diagnoses, is recorded later in the case. We thus divided the variable types into two classes. In the first class, only variables that are available at the beginning of the cases are considered. These will be used to build a minimal model to estimate the in-hospital mortality risk right at the beginning of the case. The second class contains all variable types and is used to build a full model based on all available information.

All cases have been transformed into a set of items to enable ARM. This was done by adding all variables that occurred during a case to the case's itemset. We did not exclude any case in order to get a general in-hospital mortality risk estimation model.

This results in 179,857 clinical variables and 524,520 cases, in 9,369 (1.79%) of which the patient died during their stay in the hospital. This in-hospital mortality rate is comparable to similar populations like England (2).

2.2. Rule Mining

Due to the presence of rare diseases or smaller patient subgroups, infrequent rules can be of interest in healthcare. This is why we do not use the classical ARM metrics like support and confidence (11) that focus on frequency, but instead epidemiological metrics

TABLE 2 | A fictional contingency table to show how the OR is calculated.

	Died	Survived
Items in *A* occurred	$a = 100$	$b = 200$
Items in *A* did not occur	$c = 300$	$d = 400$

that are widely used to measure associations between clinical variables and outcomes. We use the odds ratio (OR) as the primary metric to measure this association. Odds ratios have previously been used in ARM in healthcare (13). Starting from a contingency table like the one in **Table 2**, the OR can be calculated as

$$\mathrm{OR} = \frac{ad}{bc} = \frac{100 \cdot 400}{200 \cdot 300} = \frac{4}{6} = 0.\bar{6}, \tag{1}$$

indicating that there is a negative association between the items in *A* and in-hospital mortality. The set *A* can contain one or more items of one or more different types. This flexibility makes ARM techniques easy to use with large, heterogeneous data. ORs range from 0 to $+\infty$. An OR of 1.0 denotes no classification, while an $\mathrm{OR} \leq 1.0$ denotes a positive or negative association, respectively. The higher (in the case $\mathrm{OR} > 1.0$) or lower ($\mathrm{OR} < 1.0$), the stronger the association.

The mining process consists of two steps. In the first step, a contingency table like the one in **Table 2** is constructed for each variable by counting the cases with and without the variable as well as with and without in-hospital mortality. From this table, the odds ratio is calculated according to Equation (1). Note that this also includes ORs lower than 1.0, which denotes a negative association. ORs of 0 or ∞ are discarded. This corresponds to one of the cell entries a, b, c and d being 0.

In the second step, the model size is reduced by applying a filter to the rules constructed in the first step. As an OR of 1.0 denotes no association, we use a statistical hypothesis test to ensure that the calculated OR differs significantly from 1.0. The normal approximation of the log odds ratio (29) is used. The null hypothesis is "$\mathrm{OR} = 1.0$," and the test returns a two-sided p-value. Only ORs with a p-value below a configurable value $p_{\max}$ are kept. As this method results in many tests on the same dataset, Bonferroni correction (30) can be used. This procedure divides $p_{\max}$ by the number of tests to be executed and uses this quotient as the threshold value instead of $p_{\max}$. All rules of the form "variable $\Rightarrow$ in-hospital mortality," together with their OR, that are left after this filtering step then form the prediction model.

2.3. Prediction

Given a new observation, the prediction model can be used to estimate the corresponding patient's in-hospital mortality risk using the model's rules. First, all the rules that apply to the model are determined. The remaining rules do not play a role in this observation's prediction. The ORs of these applying rules are then aggregated by calculating their average value. A decision boundary δ is used to decide which average OR leads to the prediction of high in-hospital mortality risk. If OR_p is the average OR of all rules that apply to observation p, then the prediction model's decision function $f(p)$ is

$$f(p) = \begin{cases} \text{high in-hospital mortality risk,} & \text{if } \mathrm{OR}_p \geq \delta, \\ \text{low in-hospital mortality risk,} & \text{if } \mathrm{OR}_p < \delta. \end{cases} \tag{2}$$

2.4. Experiments

The following hyperparameters were used. The p-value threshold $p_{\max}$ decides which rules are kept in the filtering step. We used the thresholds 10^{-n} for n in $\{0, 1, \ldots, 10\}$ as well as the commonly used value 0.05. All of these thresholds were used with and without Bonferroni correction. Additionally, a value of 0 was used to only keep ORs with a p-value of exactly 0. This was possible due to the high number of observations. As computers use finite representations of floating-point numbers, this should be understood to be exactly zero after internal rounding. This resulted in $13 \cdot 2 = 26$ experiments. Each experiment was repeated ten times. Each time, the dataset was randomly split into 90% training set and 10% test set. The primary metric of interest was the area under the receiver operating characteristic curve (AUC) (31). The AUC values were calculated using the Accord Framework, version 3.8.0 (32). All metrics have been calculated both on the training and the test set to analyze possible overfitting. The experiments were executed for both the full model and the minimal model.

3. RESULTS

The AUC values achieved by the proposed method can be seen in **Figure 2**. For the full model, the mean AUC values on the test set range from 0.977 to 0.980, which shows that the filtering step has a minor impact on the predictive performance. With a range of 0.687 to 0.690, this is also true for the minimal model. In both cases, the difference to the training set is low, indicating that the method does not suffer from overfitting. The largest deviation in mean AUC values was 0.001 for the full model, $p_{\max} = 1$ and no Bonferroni correction.

Figure 3 shows the number of rules in both models. In the minimal model, the number of rules is almost constant. This can be explained by the small number of variables, as only 20 variables are considered. In the full model, on the other hand, almost 180,000 variables can be included in the model. The number of rules grows exponentially, with Bonferroni correction slowing the growth down considerably. Still, the number of rules reaches into the thousands. As such a large number of rules is hard to handle, a strict p-value filter is advisable if interpretability is of interest. As the performance stays almost the same with lower $p_{\max}$ values, but the number of rules is considerably lower, we can deduce that the statistical significance test is an effective filter that greatly reduces the size of the prediction model without compromising the predictive performance. Still, around 1,220 rules remain even for the full model with $p_{\max} = 0$. As can be seen in **Figure 4**, this is due to the complexity of the problem. In-hospital mortality can be caused and influenced by many factors. Per observation, however, only around 22 to 36 rules apply when using the full model, depending on $p_{\max}$ and the use

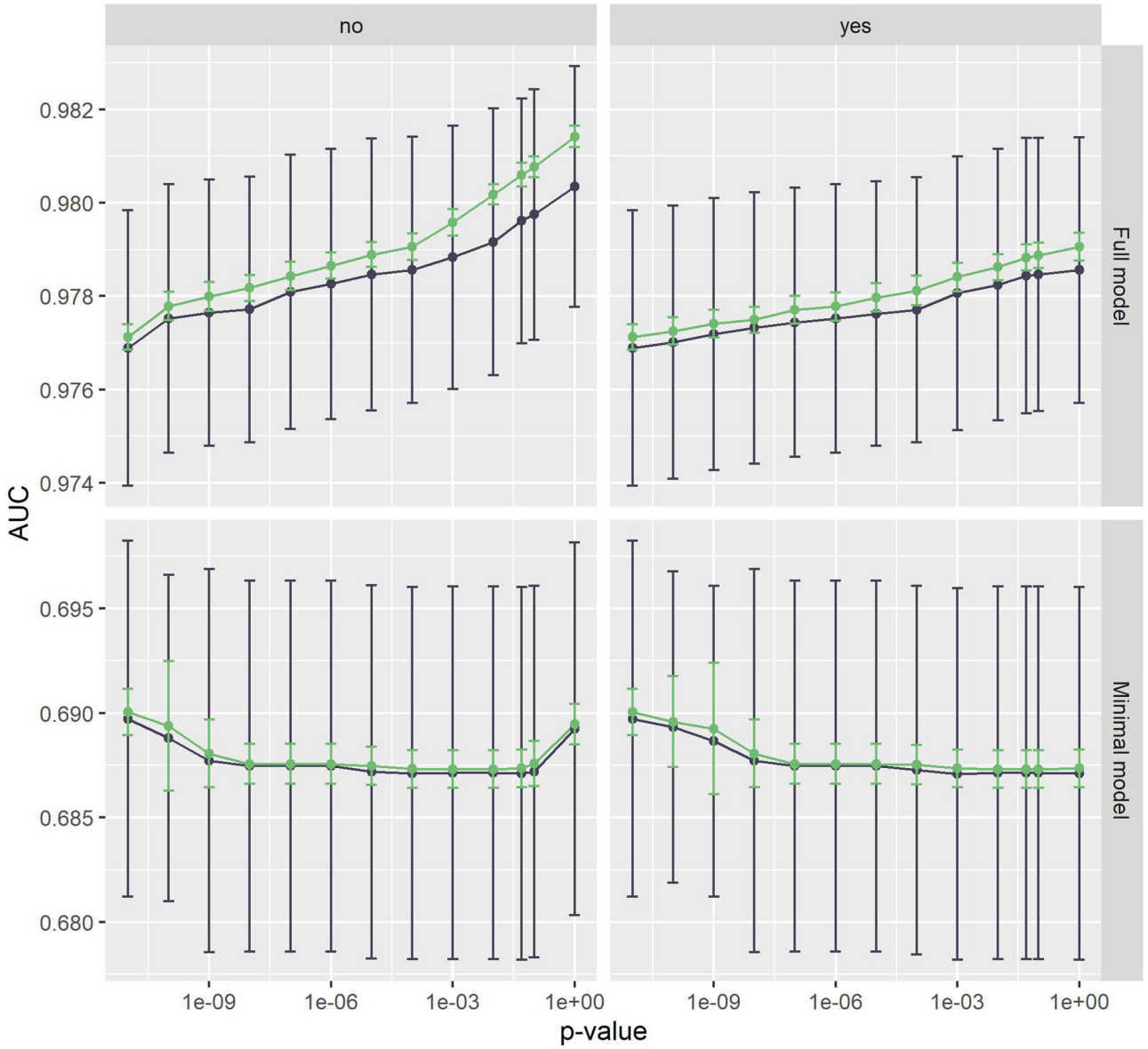


FIGURE 2 | The AUC values of the full model (top) and the minimal model (bottom). Error bars show one standard deviation, the color indicates whether the test or the training set was used. The left panel is without, the right panel with Bonferroni correction. The x-axis shows different *p*-values used in the filtering step. The leftmost dot denotes the *p*-value zero. The full model clearly outperforms the minimal model. Both models are very stable with respect to both the *p*-value threshold and Bonferroni correction.

of Bonferroni correction. On average, around 4 to 5 rules apply when using minimal model.

In the remainder of this study, we further analyze small prediction models, as they are easier to manage and interpret and have a comparable predictive performance. For both the full model and the minimal model, the lowest threshold $p_{\max} = 0$ was used. In this case, Bonferroni correction makes no difference. This results in a full model with 1,217 rules and an AUC of 0.98 and a minimal model with 13 rules and an AUC of 0.69.

The full model clearly outperforms the minimal model. This is to be expected, as the minimal model only contains very limited data. The variable types ethnicity, gender, insurance type, language, and marital status contain no information that could

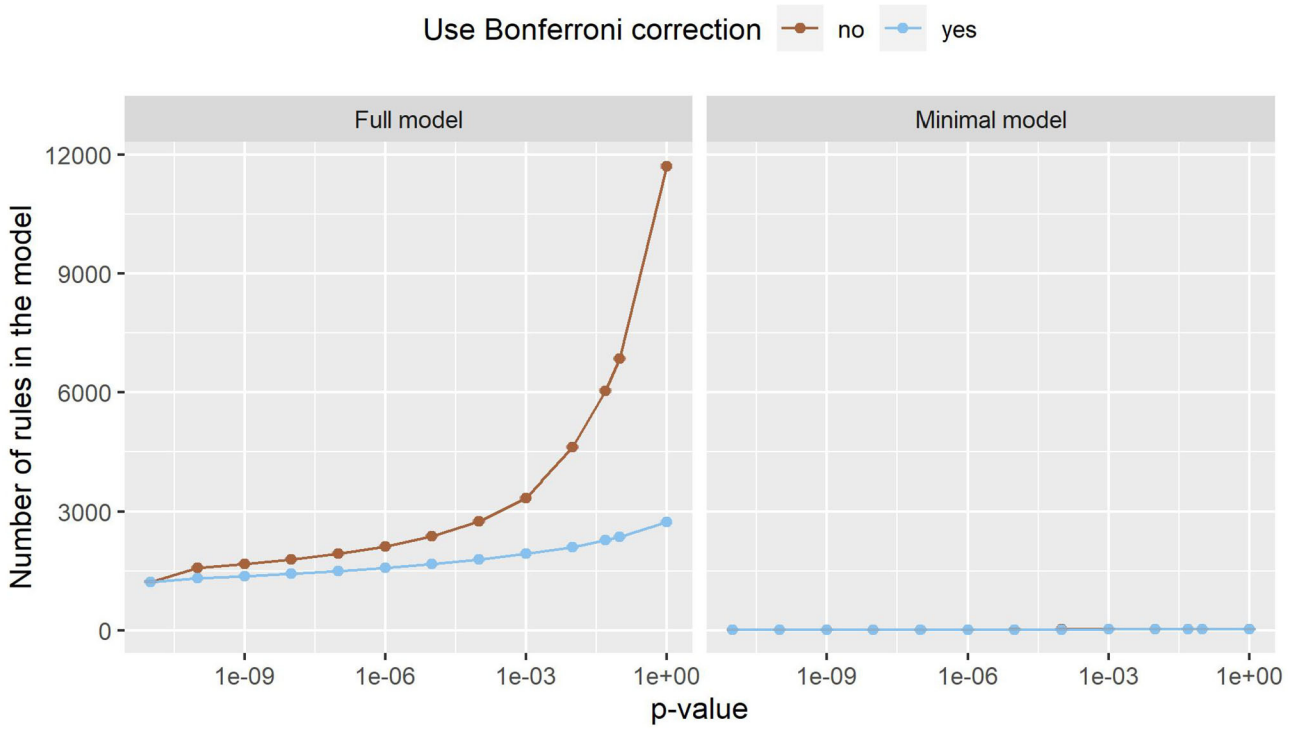


FIGURE 3 | The number of rules the full model (left) and the minimal model (right). Error bars (which are very small) show one standard deviation, the color indicates whether Bonferroni correction was used. The x-axis shows different *p*-values used in the filtering step. The leftmost dot denotes the *p*-value zero. The number of rules increases exponentially in the full model, while this effect is much smaller in the minimal model. The minimal model has fewer rules than the full model.

help determine the cause of the clinical stay or the patient's health status. However, it is noteworthy that the minimal model still achieves an AUC of 0.69 with this limited information. As all this information is readily available and patient-reported, the model can be used as a first assessment before any provider encounter. The corresponding receiver operating characteristic (ROC) curve can be seen in **Figure 5**. To choose a decision boundary, we searched for the highest Youden index (33), which is equivalent to optimizing the sum of sensitivity and specificity. The corresponding minimal model uses a decision boundary of 1.015, which results in an accuracy of 63%, a sensitivity of 66%, and a specificity of 63%. Other decision boundaries are possible, depending on the context in which they are used. Varying the decision boundary will affect both sensitivity and specificity of the model.

This additional information in the full model improves the predictive performance considerably. With an AUC of 0.98, the model can almost perfectly predict the in-hospital mortality risk. The corresponding ROC curve can be seen in **Figure 6**. With an optimal decision boundary of 5.306, this results in an accuracy of 93%, a sensitivity of 95%, and a specificity of 93%.

All analyses were done in the R programming language (34), using the tidyverse packages (35). The rules of both models can be found in the **Supplementary Materials**.

3.1. Analysis of Rules

Apart from their predictive qualities, the rules in the models also allow us to analyze the algorithm's reasoning. The 13 rules in the minimal model span all five included variable types. These rules indicate that male patients die more often than female patients (ORs 1.25 vs. 0.80), that English speakers are more likely to survive than others (0.73 vs. 1.36), and that Medicare patients are more likely to die than Medicaid and other patients (2.40 vs. 0.57 vs. 0.50). The ethnicities 'Hispanic/Latino' and 'Black/African American' are negatively associated with in-hospital mortality (0.47 and 0.58), while 'unknown' and 'unable to obtain' have higher ORs (4.59 and 2.25). The marital status 'widowed' is positively associated with in-hospital mortality (2.05), while being single is negatively associated with in-hospital mortality (0.50).

From the five variable types in the minimal model, the same 13 rules are in the full model. The remaining 1,204 rules span over the five other variable types with 602 diagnosis, 445 prescription, 131 procedure, nine service, and 17 ward rules.

Only five diagnosis rules are negative associations, namely three diagnoses on single liveborn infants (0.08-0.16), encounters for immunizations (0.06), and unspecified chest pain (0.11). The remaining 597 rules can be categorized into various groups of rules that describe variants of the same pattern.

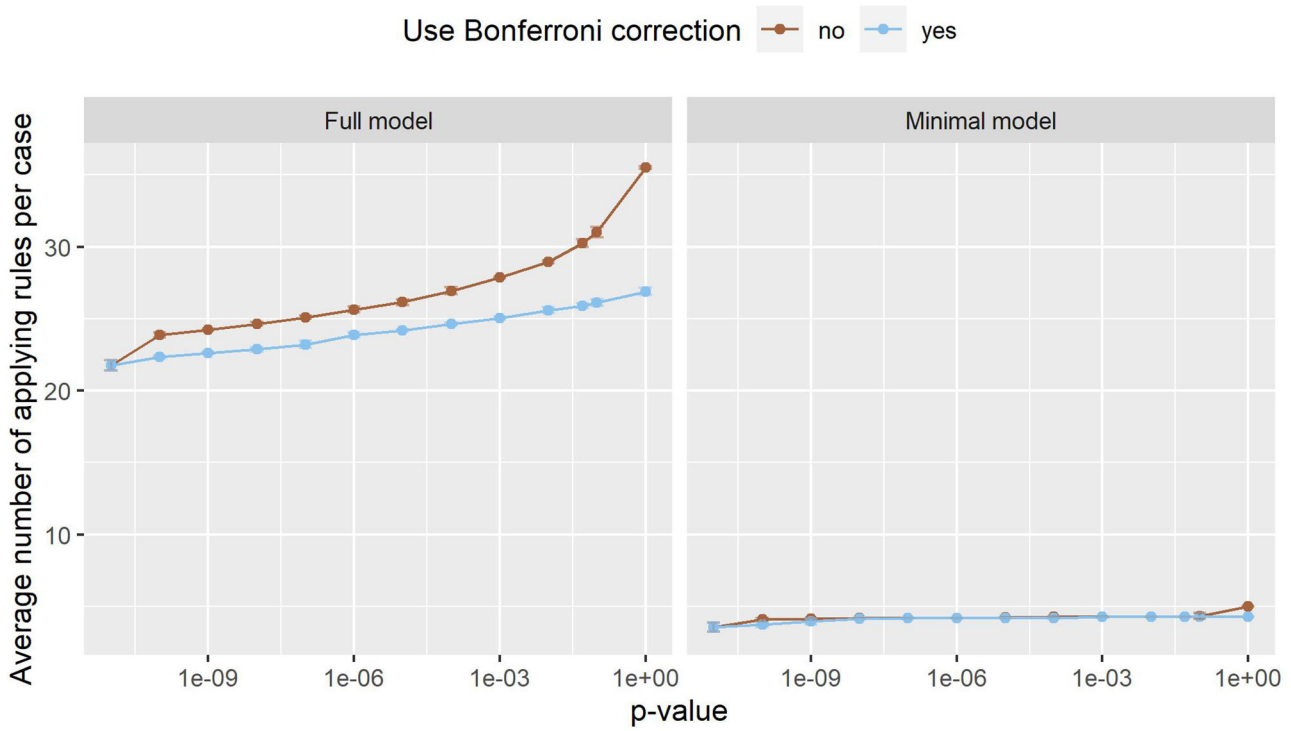


FIGURE 4 | The number of rules that apply to an average case in the full model (left) and the minimal model (right). Error bars (which are very small) show one standard deviation, the color indicates whether Bonferroni correction was used. The x-axis shows different *p*-values used in the filtering step. The leftmost dot denotes the *p*-value zero. Only a small fraction of all rules apply to an individual case, making it easy to interpret the model and its decisions.

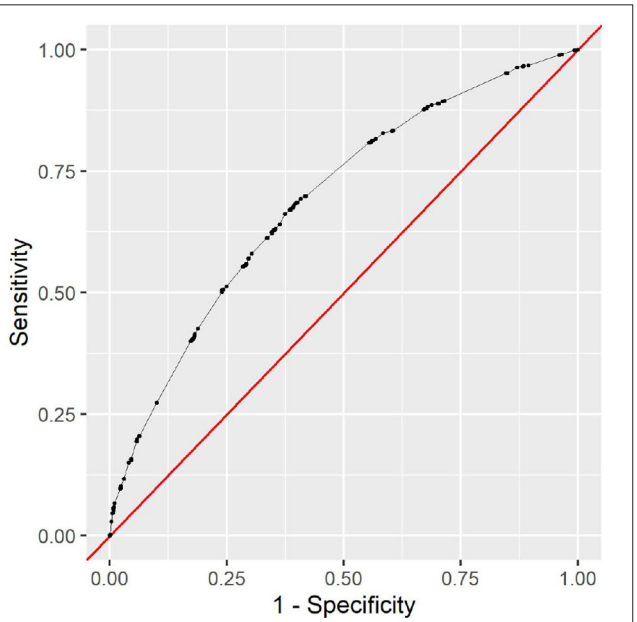


FIGURE 5 | The receiver operating characteristic of the minimal model.

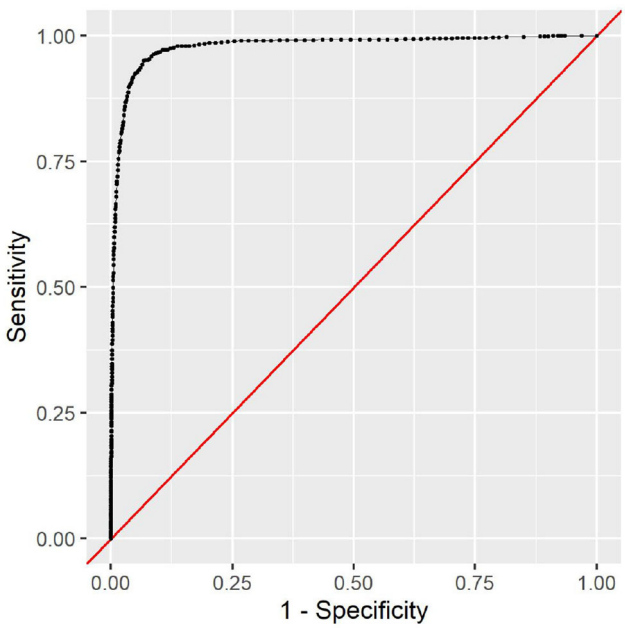


FIGURE 6 | The receiver operating characteristic of the full model.

Examples include alcohol abuse and its consequences (2.34–11.69), anemia (1.67–4.56), various forms of hemorrhage (3.32–582.13), neoplasms (3.50–13.66), pneumonia (6.27–18.49), pressure ulcers (5.29–17.96), sepsis and septicemia (4.15–33.18), and diabetes (1.56–13.20). Some doubling occurs due to two

TABLE 3 | A list of rules that apply to a fictional patient.

Variable type	Description	OR	Minimal model?
Ethnicity	Black/African American	0.58	✓
Gender	Female	0.80	✓
Insurance type	Medicaid	0.57	✓
Language	English	0.73	✓
Marital status	Single	0.50	✓
Diagnosis	Acidosis	10.96	
Diagnosis	Anuria and oliguria	17.27	
Prescription	Sodium Bicarbonate	11.54	
Prescription	Furosemide	5.00	

This is also how the applying rules could be shown to providers.

versions of the International Classification of Diseases being used in the dataset.

This also concerns the 131 procedure rules. Here, examples of groups of rules are catheterizations (2.59–21.11), drainages (5.53–22.06), ventilation (2.44–26.02), and transfusions (3.66–13.51). No procedure rule has an OR below 1.0.

Only 17 out of 445 prescription rules are negative. As some drugs are recorded with slightly different names, different doses or different mode of administration, some doubling occurs. One extreme example is sodium chloride with 20 rules. The rules "0.45% Sodium Chloride" and "0.45 % Sodium Chloride" (note the space) have ORs of 2.11 and 94.94, respectively. The difference in ORs can be explained by the unequal distribution across units. While 18 of the 23 (78%) cases with "0.45 % Sodium Chloride" in MIMIC-IV were in at least one Intensive Care Unit, only 3375 of the 15,180 (22%) cases with "0.45% Sodium Chloride" were in at least one Intensive Care Unit. This indicates that there are some unexpected inconsistencies in the data. With almost 180,000 variables, it would be impossible to check all variables manually for inconsistencies due to variations in practice. Other prescription rule groups with high variation include Heparin (2.10–68.58), vaccines (0.01–2.82), and lidocaine (2.10–15.55).

Three service rules are negative, namely obstetrics (0.01), newborn (0.17), and orthopedic (0.19). The positive rules include four medical services (1.64–2.62), trauma (2.39), and neurologic surgical (2.73).

Finally, five out of 17 ward rules are negative, with three being at least partially about childbirth (0.02–0.48). The six rules with the highest ORs (7.55–13.25) are all the intensive care units in the dataset. Both the service and ward rules reflect different patient populations with different reasons for the clinical stay. The in-hospital mortality risk is vastly different between pregnant women and traffic accident victims that need intensive care, for example.

Overall, only 37 of 1,217 rules are negative. The five rules with the lowest ORs are about childbirth (0.00–0.04) and a Hepatitis B vaccine (0.01), the five highest ORs are medications used in palliative care (Morphine and Angiotensin II, 856.26 and 332.77), subdural hemorrhage (582.14), and two rules on brain death (413.38 and 668.41). These last two rules are unexpected. Brain death should always co-occur with in-hospital mortality, which would exclude this OR of ∞. Their occurrence hints at an error in the data. Braindead patients who are organ donors are recorded as having died, but then another case is opened for them with the diagnosis of brain death, but it is recorded that they survived this second case. This leads to more inconsistent data contained in MIMIC-IV, which can now be seen in the prediction model.

3.2. Interpretability

To study the proposed method's interpretability, we give a fictional example of how the prediction model can be used. **Table 3** shows the rules that apply to a fictional patient in the emergency department. This list of rules is how the prediction model's decision could be shown to providers. Apart from the decision, the rules also explain what the case is about. The items in the minimal model give us a first assessment. The patient is female, Black/African American, single, insured under Medicaid, and speaks English. This results in an average OR of 0.64, which is below the decision boundary of 1.015. This tells us that, based on the minimal model, our patient is low-risk.

In the full model, more information becomes available. For example, suppose that the following information has become available after 20 min in the emergency department. We now know that acidosis and anuria or oliguria was diagnosed. The patient was given Sodium Bicarbonate and Furosemide and is currently in the emergency department. While this is not much information due to the short length of stay, it allows us to update our risk estimation. With an average OR of 5.328, which is greater than the full model's decision boundary of 5.306, we conclude that the patient currently has a high in-hospital mortality risk, but she is close to the decision boundary. We see that not only are the models interpretable, but they can also be used to get a quick overview of the case at hand.

3.3. Comparison to Decision Trees

We compared our models to Decision Trees models, which are the state of the art in interpretable machine learning for in-hospital mortality as they show the best trade-off between performance and interpretability (6).

We trained Decision Tree classifiers using scikit-learn, version 0.24.2 (36), with varying maximal depths. As the dataset is highly

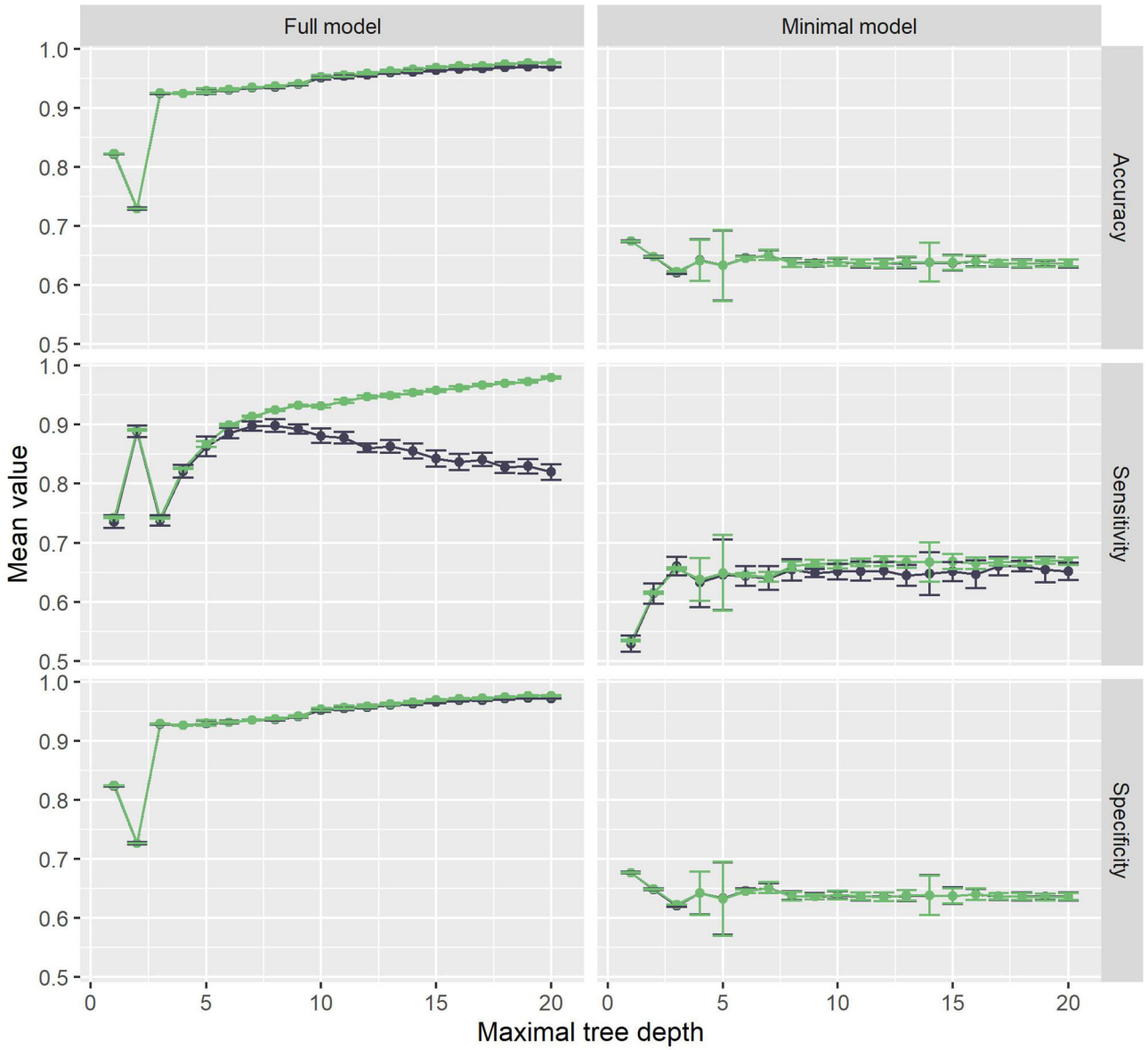


FIGURE 7 | Various metrics on the Decision Tree models. The left panels are the full model, while the right panels are the minimal model. The color denotes whether the training or the test set was used to produce these metrics. On the x-axis, various maximal tree depths were analyzed. For the full model, there is major overfitting. The test set sensitivity quickly drop below the training set sensitivity for higher maximal tree depths. For the minimal model, the sensitivity also drops below the training set's sensitivity.

imbalanced, with over 98% of patients surviving, we used the balanced class weight option to give more weight to the rarer class. These experiments were executed for both the full and the minimal model.

The results are visualized in **Figure 7**. For the minimal model, the results are very similar to our proposed method. Above a maximal tree depth of five, the metrics are stable, with only small deviations in the sensitivities of the training and test set.

Figure 8 shows the number of leaves in the trees, which can be understood as the number of possible paths in the tree and thus as the number of rules that can be extracted from such a tree. The number of leaves in the minimal Decision Tree model grows quickly to around 433. The growth stops at a maximal tree depth of around 15, and the model's size stays relatively stable. A maximal depth of at least six is needed to achieve a stable model with low standard deviations and satisfactory metrics, at which

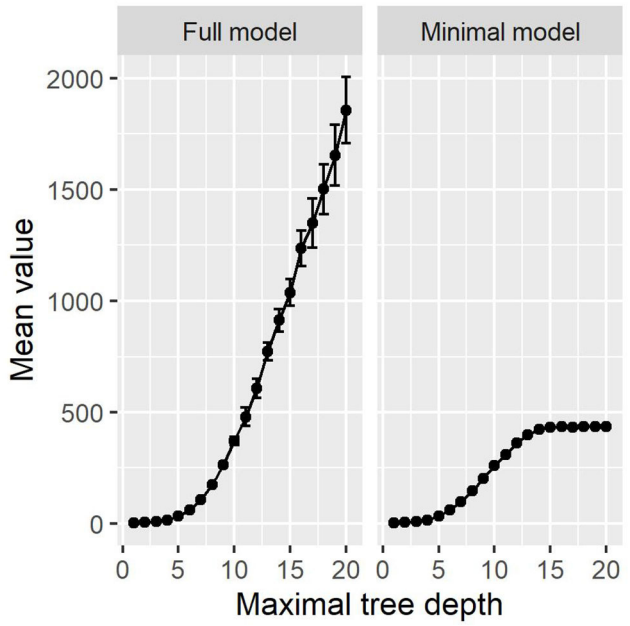


FIGURE 8 | The number of leaves in the Decision Tree models. The full model increases exponentially. The minimal model's size is constant above a depth of around 15.

point the model already contains around 60 leaves. Our proposed method's minimal model only needs 13 rules for comparable performance. We thus see that while the predictive performance is similar to our proposed minimal model, the complexity of the model is higher.

The full Decision Tree model suffers from major overfitting, as can be seen in the sensitivity values of the test set, which drop below the training set sensitivity for maximal tree depth values above five. With the highest sensitivity of 90%, Decision Trees cannot outperform our proposed method's full model in terms of detecting high-risk cases. On the other hand, decision Trees achieve a slightly higher specificity of 97% and thus detect low-risk cases more reliably. Note that the choice of a decision boundary impacts both sensitivity and specificity of the proposed method, and another decision boundary for the full model (namely 6.081) leads to almost the same metrics as the full Decision Tree model (namely a sensitivity of 90% and a specificity of 96%). In this way, the higher specificity at the cost of a lower sensitivity can also be achieved with the full model.

As can be seen in **Figure 8**, the number of leaves grows exponentially in the full model, just as the number of rules does. In comparison to **Figure 3**, there are more rules in our proposed method then leaves in the Decision Tree methods. With Decision Trees, however, the paths are much more complex, as their length can go up to the maximal tree depth. We provide one example full Decision Tree model with a maximal tree depth of seven in the **Supplementary Materials**. At this depth, the Decision Tree model does not yet suffer from major overfitting, resulting in a test accuracy of 93%, sensitivity of 90%, and specificity of 94%. The tree contains 103 leaves and 204 decision nodes.

Following the first chain in the tree, the model decides that the patient is low-risk due to the variables "5% Dextrose," "Morphine Sulfate," "Encounter for palliative care," "Insertion of endotracheal tube," "Encounter for palliative care," and "Vasopressin" being absent in the case. If the diagnosis "Less than 24 completed weeks of gestation" were present, the model would decide that the patient is high-risk. This means that the variables' influence is combined into a chain of yes-no choices, and one change in the decision chain can alter the model's decision. This hides each variable's influence on the decision as it could occur multiple times in the tree, but each patient only activates one chain. Our proposed method separates all the variables into one-to-one rules, making it easier to identify each variable's importance individually. The combination of the variables into one decision takes place during the prediction, where the whole context of the case is taken into account and we get an overview of the case.

Another challenge in the interpretation of decision tree models is the combination of variables in each chain. The variable "Encounter for palliative care" occurs twice in the same chain due to two International Classification of Diseases versions being used. Additionally, the variable "Less than 24 completed weeks of gestation" does not match the palliative care diagnoses. The tree structure does not allow us to get an overview of the case at hand as it also highlights variables that did not occur during the case, obfuscating what actually happened. Our full model shows which rules apply to the case and thus give us an explanation of what happened during the case. Building the decision on top of this knowledge helps providers understand the model's decision.

In summary, Decision Tree models do not offer performance improvements and show limited interpretability compared to our proposed model. This is due to the fact that more complexity in the Decision Tree Models is needed to achieve a comparable predictive performance and because Decision Trees combine various variables in chains of yes-no-choices. Additionally, Decision Tree models introduce order to the variables, as the tree-like structure divides the dataset sequentially. On the other hand, our model treats every variable individually, independent of the other variables, making it easier to analyze each variable's effect.

4. DISCUSSION

Associative classification based on ORs is a feasible method for in-hospital mortality risk estimation. Apart from the high predictive quality of the full model, the proposed method also allowed us to analyze the dataset based on the rules contained in the prediction model. The proposed method is not prone to overfitting and generalizes well.

While the AUCs of the minimal model are much lower, its predictive qualities are relatively high. Only socio-demographic information that can be provided by the patient is added, there is no information on the case at hand. This hints at underlying structural differences. The information what its effects are on individual patients can be valuable information for providers.

The fictional patient described above is an example of how the prediction models can be easily explained to providers. This

interpretability also helps us understand the clinical data and discover patterns and inconsistencies in the data. As shown in the examples of brain death and sodium chloride, inconsistencies in the data show up as unexpected rules or rules with similar variables but vastly different ORs.

Other rules can serve as starting points for future studies. One example are the language rules. The evidence in support of an effect of primary language on in-hospital mortality is limited (37, 38). Nevertheless, our rules suggest some underlying effect exists with ORs of 0.73 for English speakers and 1.36 for others. Further research is needed to analyze the differences in these groups and whether this effect is due to documentation practice.

The majority of rules reproduce known associations or associations that are evident and explainable with expert knowledge. Examples for the latter include the higher mortality in intensive care units due to the higher number of critical cases, the low mortality in childbirth rules, the high mortality in Medicare patients due to confounding by age, and various diagnoses that are either indicative of palliative care (like Morphine or Angiotensin II) or of low-risk cases like immunizations.

Some of the rules for which previous research exists include the following. Diabetes insipidus (13.20) has previously been identified as a potential cause of missed care and increased mortality (39, 40). Alkalosis (5.39–8.10) is associated with increased mortality (41). Alcohol dependence (2.34-11.69) can lead to various health problems of cognitive, cardiovascular, and gastrointestinal nature, among others (42). Each of these, in turn, can contribute to increased mortality and emergencies that lead to increased in-hospital mortality.

4.1. Limitations and Outlook

The present study has major limitations. First, only categorical data that changes slowly is considered. This excludes other interesting data like vital parameters and many biomarkers. Due to the rule-based approach, quantitative data have to be grouped into bins. Quickly changing data could easily be added to the method, but this would require analysis on a higher temporal resolution. Timestamps for all of the variables in the underlying dataset would help get more information from each case.

A second limitation is the lack of causal explanations. All the rules in both models are correlations between the variable and in-hospital mortality. They do not explain why this correlation exists. Future work is needed to introduce causal inference mechanisms into the presented approach.

There are several further possibilities for future work.

While the inconsistencies encountered in the dataset do not harm the proposed method's predictive performance, it might be helpful to remove them. However, it requires efforts to fix potential inconsistencies in thousands of clinical variables. Future research is needed to assess the consequences of resolving the inconsistencies as well as the potential for automated solutions to do so.

Longer rules, i.e., rules with more than one item on the left-hand side, can be created using ORs. This has not been analyzed in this study for two reasons. First, the predictive qualities are very good as-is, so more complex rules are not expected to bring much improvement. Second, more complex rules hinder interpretability. In the present form, the rules separate all variables and make them analyzable in isolation. This sets the proposed approach apart from Decision Trees, where many variables are merged into more complex rules, lowering the method's interpretability.

The proposed method was tested and validated using in-hospital mortality as an example. Other variables in MIMIC-IV could be used as the outcome of interest without changes to the model. This, of course, results in other sets of rules and different predictive performance, but the interpretable nature of the model remains the same.

One open question is the usability of the proposed method in clinics and hospitals. A usability study could be used to assess whether the proposed method is helpful to providers in realistic scenarios and whether it can replace or complement existing manual scoring methods. This is expected to highlight potential challenges in and benefits from the implementation of the proposed method.

5. CONCLUSION

We proposed a novel associative classification method to estimate a patient's in-hospital mortality risk. With a minimal model that uses patient-reported information that is quickly available and a full model that uses more information that is available at a later point in time, providers can objectively track a patient's in-hospital mortality risk. Apart from the high predictive performance, the rule-based nature of the proposed method allowed us to analyze which among around 180,000 variables play a role in the estimation of in-hospital mortality risk, resulting in a model with around 1,200 variables.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://physionet.org/content/mimiciv/0.4/.

AUTHOR CONTRIBUTIONS

OH: concept, programming, analysis of the results, and writing of the manuscript. AM and ER: substantial revision of the work. All authors have approved the submitted version and take responsibility for the scientific integrity of the work.

FUNDING

This project was funded by the Bavarian State Ministry of Science and the Arts, coordinated by the Bavarian Research Institute for Digital Transformation (bidt), and supported by the Bavarian Academic Forum (BayWISS)–Doctoral Consortium Health Research.

ACKNOWLEDGMENTS

The authors wish to thank Felix Meister for the valuable discussions that have greatly improved the present study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.785711/full#supplementary-material

Supplementary Table 1 | The rules in the full model.

Supplementary Table 2 | The rules in the minimal model.

Supplementary Table 3 | One full Decision Tree model.

REFERENCES

1. Wieczorek-Wojcik B, Gaworska-Krzemińska A, Owczarek AJ, Kilańska D. In-hospital mortality as the side effect of missed care. *J Nurs Manag*. (2020) 28:2240–6. doi: 10.1111/jonm.12965
2. Stewart K, Choudry MI, Buckingham R. Learning from hospital mortality. *Clin Med*. (2016) 16:530. doi: 10.7861/clinmedicine.16-6-530
3. Salluh JI, Soares M. ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care*. (2014) 20:557–565. doi: 10.1097/MCC.0000000000000135
4. Bishop CM. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin; Heidelberg: Springer-Verlag (2006).
5. Johnson AEW, Pollard TJ, Mark RG. Reproducibility in critical care: a mortality prediction case study. In: Doshi-Velez F, Fackler J, Kale D, Ranganath R, Wallace B, Wiens J, editors. *Proceedings of the 2nd Machine Learning for Healthcare Conference, Vol. 68 of Proceedings of Machine Learning Research*. Boston, MA: PMLR (2017). p. 361–76.
6. Xie J, Su B, Li C, Lin K, Li H, Hu Y, et al. A review of modeling methods for predicting in-hospital mortality of patients in intensive care unit. *J Emerg Crit Care Med*. (2017) 1:e18. doi: 10.21037/jeccm.2017.08.03
7. Keuning BE, Kaufmann T, Wiersema R, Granholm A, Pettilä V, Møller MH, et al. Mortality prediction models in the adult critically ill: a scoping review. *Acta Anaesthesiol Scand*. (2020) 64:424–442. doi: 10.1111/aas.13527
8. Stiglic G, Kocbek P, Fijacko N, Zitnik M, Verbert K, Cilar L. Interpretability of machine learning-based prediction models in healthcare. *WIREs Data Min Knowl Discov*. (2020) 10:e1379. doi: 10.1002/widm.1379
9. Fu LH, Schwartz J, Moy A, Knaplund C, Kang MJ, Schnock KO, et al. Development and validation of early warning score system: a systematic literature review. *J Biomed Inform*. (2020) 105:e103410. doi: 10.1016/j.jbi.2020.103410
10. Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and Regression Trees*. Boca Raton, FL: Wadsworth International Group (1984).
11. Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases. In: *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data. SIGMOD '93*. New York, NY: Association for Computing Machinery (1993). p. 207–16.
12. Thabtah F. A review of associative classification mining. *Knowl Eng Rev*. (2007) 03;22:37–65. doi: 10.1017/S0269888907001026
13. Lin WY, Li HY, Du JW, Feng WY, Lo CF, Soo VW. iADRs: towards online adverse drug reaction analysis. *Springerplus*. (2012) 1:e72. doi: 10.1186/2193-1801-1-72
14. El Houby EM. A framework for prediction of response to HCV therapy using different data mining techniques. *Adv Bioinformatics*. (2014) 2014:e181056. doi: 10.1155/2014/181056
15. Uriarte-Arcia AV, López-Yáñez I, Yáñez-Márquez C. One-hot vector hybrid associative classifier for medical data classification. *PLoS ONE*. (2014) 9:e95715. doi: 10.1371/journal.pone.0095715
16. Kadkhoda M, Akbarzadeh-T MR, Sabahi F. FLeAC: a human-centered associative classifier using the validity concept. *IEEE Trans Cybern*. (2020). 1–12. doi: 10.1109/TCYB.2020.3025479
17. Dua S, Singh H, Thompson HW. Associative classification of mammograms using weighted rules. *Expert Syst Appl*. (2009) 36:9250–59. doi: 10.1016/j.eswa.2008.12.050
18. Rea S, Huff S. Cohort amplification: an associative classification framework for identification of disease cohorts in the electronic health record. *AMIA Ann Symp Proc*. (2010) 2010:862–6.
19. Ujager FS, Mahmood A. A context-aware accurate wellness determination (CAAWD) model for elderly people using lazy associative classification. *Sensors (Basel)*. (2019) 19:e1613. doi: 10.3390/s19071613
20. Meena K, Tayal DK, Gupta V, Fatima A. Using classification techniques for statistical analysis of Anemia. *Artif Intell Med*. (2019) 94:138–52. doi: 10.1016/j.artmed.2019.02.005
21. Kianmehr K, Alhajj R. CARSVM: a class association rule-based classification framework and its application to gene expression data. *Artif Intell Med*. (2008) 44:7–25. doi: 10.1016/j.artmed.2008.05.002
22. He Y, Hui SC. Exploring ant-based algorithms for gene expression data analysis. *Artif Intell Med*. (2009) 47:105–19. doi: 10.1016/j.artmed.2009.03.004
23. Yu P, Wild DJ. Fast rule-based bioactivity prediction using associative classification mining. *J Cheminform*. (2012) 4:e29. doi: 10.1186/1758-2946-4-29
24. Yu P, Wild DJ. Discovering associations in biomedical datasets by link-based associative classifier (LAC). *PLoS ONE*. (2012) 7:1–11. doi: 10.1371/journal.pone.0051018
25. ElHefnawi M, Sherif FF. Accurate classification and hemagglutinin amino acid signatures for influenza a virus host-origin association and subtyping. *Virology*. (2014) 449:328–38. doi: 10.1016/j.virol.2013.11.010
26. Cheng CW, Wang MD. Improving personalized clinical risk prediction based on causality-based association rules. *ACM BCB*. (2015) 2015:386–92. doi: 10.1145/2808719.2808759
27. Haas O, Maier A, Rothgang E. Using associative classification and odds ratios for in-hospital mortality risk estimation. In: *Workshop on Interpretable ML in Healthcare at International Conference on Machine Learning (ICML)*. (2021).
28. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. MIMIC-IV (version 0.4). *PhysioNet*. (2020). doi: 10.13026/a3wn-hq05
29. Morris JA, Gardner MJ. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br Med J (Clin Res Ed)*. (1988) 296:1313–6. doi: 10.1136/bmj.296.6632.1313
30. Shaffer JP. Multiple hypothesis testing. *Annu Rev Psychol*. (1995) 46:561–84. doi: 10.1146/annurev.ps.46.020195.003021
31. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett*. (2006) 27:861–74. doi: 10.1016/j.patrec.2005.10.010
32. Souza C, Kirillov A, Catalano MD, Contributors AN. *The Accord*.NET Framework (2014). Available online at: http://accord-framework.net.
33. Youden WJ. Index for rating diagnostic tests. *Cancer*. (1950) 3:32–5. doi: 10.1002/1097-014219503:1<32::AID-CNCR2820030106>3.0.CO;2-3
34. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna (2021). Available online at: https://www.R-project.org/.
35. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. *J Open Source Softw*. (2019) 4:1686. doi: 10.21105/joss.01686
36. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30.
37. John-Baptiste A, Naglie G, Tomlinson G, Alibhai S, Etchells E, Cheung A, et al. The effect of english language proficiency on length of stay and in-hospital mortality. *J Gen Intern Med*. (2004) 19:221–8. doi: 10.1111/j.1525-1497.2004.21205.x
38. Cano-Ibáñez N, Zolfaghari Y, Amezcua-Prieto C, Khan K. Physician–patient language discordance and poor health outcomes: a systematic scoping review. *Front Public Health*. (2021) 9:629041. doi: 10.3389/fpubh.2021.629041
39. Gleeson H, Bonfield A, Hackett E, Crasto W. Concerns about patient safety in patients with diabetes insipidus admitted as inpatients. *Clin Endocrinol*. (2016) 84:950–1. doi: 10.1111/cen.13028
40. Ebrahimi F, Kutz A, Wagner U, Illigens B, Siepmann T, Schuetz P, et al. Excess mortality among hospitalized patients with hypopituitarism-a population based matched cohort study. *J Clin Endocrinol Metab*. (2020) 105:dgaa517. doi: 10.1210/clinem/dgaa517

41. Anderson L, Henrich W. Alkalemia-associated morbidity and mortality in medical and surgical patients. *South Med J.* (1987) 80:729–33. doi: 10.1097/00007611-198706000-00016
42. Schuckit MA. Alcohol-use disorders. *Lancet*. (2009) 373:492–501. doi: 10.1016/S0140-6736(09)60009-X

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

CLINICAL TRIAL
published: 16 November 2021
doi: 10.3389/fmed.2021.743009

Comparison of Cricothyroid Membrane Puncture Anesthesia and Topical Anesthesia for Awake Fiberoptic Intubation: A Double-Blinded Randomized Controlled Trial

Shaocheng Wang[1,2,3†], *Chaoli Hu*[3†], *Tingting Zhang*[3], *Xuan Zhao*[3*] and *Cheng Li*[2,3,4,5*]

[1] Department of Anesthesiology, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China, [2] Department of Anesthesiology and Perioperative Medicine, Shanghai Fourth People's Hospital, School of Medicine, Tongji University, Shanghai, China, [3] Department of Anesthesiology, Shanghai Tenth People's Hospital, School of Medicine, Tongji University, Shanghai, China, [4] Translational Research Institute of Brain and Brain-Like Intelligence, Shanghai Fourth People's Hospital, School of Medicine, Tongji University, Shanghai, China, [5] Clinical Research Center for Anesthesiology and Perioperative Medicine, Tongji University, Shanghai, China

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Jesus Rico-Feijoo,
Hospital Universitario Río Hortega, Spain
Massimiliano Sorbello,
Gaspare Rodolico Hospital, Italy

***Correspondence:**
Cheng Li
chengli_2017@tongji.edu.cn
Xuan Zhao
594079127@qq.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: 17 July 2021
Accepted: 12 October 2021
Published: 16 November 2021

Citation:
Wang S, Hu C, Zhang T, Zhao X and Li C (2021) Comparison of Cricothyroid Membrane Puncture Anesthesia and Topical Anesthesia for Awake Fiberoptic Intubation: A Double-Blinded Randomized Controlled Trial. Front. Med. 8:743009. doi: 10.3389/fmed.2021.743009

Background: Awake fiberoptic intubation (AFOI) is commonly used for patients with a difficult airway. The purpose of this study was to evaluate the efficacy of cricothyroid membrane puncture anesthesia and topical anesthesia during AFOI.

Methods: A total of 70 patients (the American Society of Anesthesiologists score I-III) with anticipated difficult airways scheduled for nonemergency surgery with AFOI were randomly slated to receive cricothyroid membrane puncture anesthesia ($n = 35$) or topical anesthesia ($n = 35$). Each group received dexmedetomidine at a dose of 1.0 μg/kg and sufentanil at a dose of 0.2 μg/kg over 10 min for conscious sedation before intubation. The endoscopy intubation, post-intubation condition, and endoscopy tolerance as scored by the anesthetists were observed. The satisfaction of the operator regarding the procedure and the satisfaction of the patient 24 h after the surgery were also recorded. We recorded the success rate of the first intubation, intubation time, and hemodynamic changes during the procedure and also the adverse events.

Results: Better intubation scores, operator satisfaction, and satisfaction of the patient were observed in the cricothyroid membrane puncture anesthesia group than in the topical anesthesia group ($p < 0.05$). The intubation time in the cricothyroid membrane puncture anesthesia group was less than that in the topical anesthesia group ($p < 0.05$). There were no significant differences in the patient tolerance scores, the success rate of the first intubation, hemodynamic changes, and adverse events between both the groups.

Conclusion: Compared with topical anesthesia, cricothyroid membrane puncture anesthesia provided better intubation conditions and less intubation time with greater satisfaction of the patient and operator during endoscopic intubation.

Clinical Trial Registration: URL: http://www.chictr.org.cn/showproj.aspx?proj=42636, Identifier: ChiCTR 1900025820.

Keywords: cricothyroid membrane puncture anesthesia, topical anesthesia, difficult airway, awake fiberoptic intubation (AFOI), dexmedetomidine (DEX), sufentanil

INTRODUCTION

The incidence of the difficult airway during clinical anesthesia is as high as 4.5–7.5% (1); this is a significant issue, as failure to maintain an unobstructed patient airway may lead to hypoxemia, brain damage, or even death (2). Awake fiberoptic intubation (AFOI) is an effective technique for the patients with difficult airways; it is considered the gold standard among intubation techniques (3, 4). Optimal intubation conditions for AFOI are as follows: the patient should be comfortable, cooperative, and have hemodynamic stability; moreover, the anesthesiologist must be able to maintain the airway of the patient with spontaneous ventilation (5). To achieve these conditions, adequate conscious sedation and high-grade local anesthesia are required. Previous studies have demonstrated that sufentanil and dexmedetomidine provide effective sedation during AFOI without depressing respiratory function (6–11).

The two most commonly used local anesthesia techniques are: cricothyroid membrane puncture anesthesia and topical anesthesia that can provide reasonable levels of safety and comfort (12–15). The duration of AFOI should be kept as short as possible to minimize the patient discomfort. Compared with the topical anesthesia performed by using the spray-as-you-go technique (16, 17), cricothyroid membrane puncture anesthesia seems to be faster and more effective (18–21). To the best of our knowledge, there is no previous study that has compared topical anesthesia and cricothyroid membrane puncture anesthesia in awake fiberoptic nasotracheal intubation. Therefore, this study was designed to compare the efficacy of the topical anesthesia and cricothyroid membrane puncture anesthesia in patients with difficult airway during AFOI.

MATERIALS AND METHODS

This study was approved by the Institutional Review Board of the Ethics Committee of Shanghai Tenth People's Hospital affiliated with the Shanghai Tongji University School of Medicine (SHSY-IEC-4.0/19-81-01). Written informed consent was obtained from each patient. It was registered as a clinical trial (www.chictr.org.cn, Identifier: ChiCTR 1900025820). We recruited 81 patients who were 18–80 years old and scheduled for AFOI due to an anticipated difficult airway with an American Society of Anesthesiologists (ASA) score of I-III. The exclusion criteria included the heart rate (HR) < 50 beats/min, systolic blood pressure (SBP) < 90 mm Hg, use of an α2-adrenoreceptor agonist or antagonist within the past 14 days, cirrhosis, nasal injury, nasal polyps, upper airway obstruction, skull base fracture, sinusitis, intracranial hypertension, heart failure, emergency surgery, coagulation disorders, contraindication to the performance of cricothyroid membrane puncture (thyroid swelling, local infection, or laryngeal disorder), allergic to related drugs and materials, cannot cooperate actively, and cannot objectively describe the symptoms. A total of 11 patients were excluded, seven patients met the exclusion criteria and four patients declined to participate. The remaining 70 patients were assigned (using a computer-generated randomization schedule) to receive topical anesthesia (Group A) or cricothyroid membrane puncture anesthesia (Group B). An anesthetist nurse generated the allocation sequence and assigned the patients to their groups, while another anesthetist nurse recorded the experimental and postoperative follow-up data. One anesthetist prepared the drug infusion, while another anesthetist was in charge of cricothyroid membrane puncture and intubation, graded intubation condition, and operator satisfaction. The patients, anesthetist nurses, and intubating anesthetist were all blinded to the group allocation.

Once the patient was transferred to the operation room, intravenous access was established and standard monitoring parameters (non-invasive blood pressure, pulse oximetry, and ECG) were recorded every 2 min. The patient inhaled oxygen through a nasal catheter (4 l/min). Topical anesthesia of the nasal cavity was initiated using 2 ml 2% lidocaine; simultaneously, 2 ml 1% ephedrine was instilled into the nasal cavity to contract the nasal vessels. Sufentanil (0.2 μg/kg) and dexmedetomidine (1 μg/kg) were diluted into 100 ml 0.9% saline and the patient received the drugs intravenously over 10 min. When drug infusion was completed, cricothyroid membrane puncture was performed by using a 23G needle. After verification of intratracheal placement by performing air aspiration, 3 ml of 2% lidocaine was injected in Group B, while 3 ml of 0.9% saline was injected in Group A. After injection, the patient was asked to cough to transport the local anesthetic from the tracheal injection site to the supraglottic mucosa. After a cricothyroid membrane puncture, a 30 sec wait was conducted as part of the protocol. When cricothyroid membrane puncture was accomplished, a fiberoptic scope (Olympus LF-DP 3.1 mm, Olympus, Tokyo, Japan) was loaded with a 7.0-mm tracheal tube for male patients or a 6.5-mm tube for female patients. Then, 2 ml of 2% lidocaine for Group A or 2 ml of 0.9% saline for Group B was sprayed directly onto the glottis through the channel of the fiberoptic scope once the glottic structures were identified. After a 1-min wait, another 2 ml of 2% lidocaine for Group A or 2 ml of 0.9% saline for Group B was sprayed below the vocal cords. After a further 1-min wait, the tracheal tube was slightly inserted *via* the fiberoptic scope tube. During intubation, if the peripheral oxygen saturation (SpO_2) of the patient fell $\leq 92\%$, the procedure was halted and the patient was asked to take deep breaths. Another intubation attempt was made when the SpO_2 was recovered to $\geq 95\%$. Hemodynamic changes (HR, mean arterial pressure, and pulse oximetry) were analyzed for both the groups

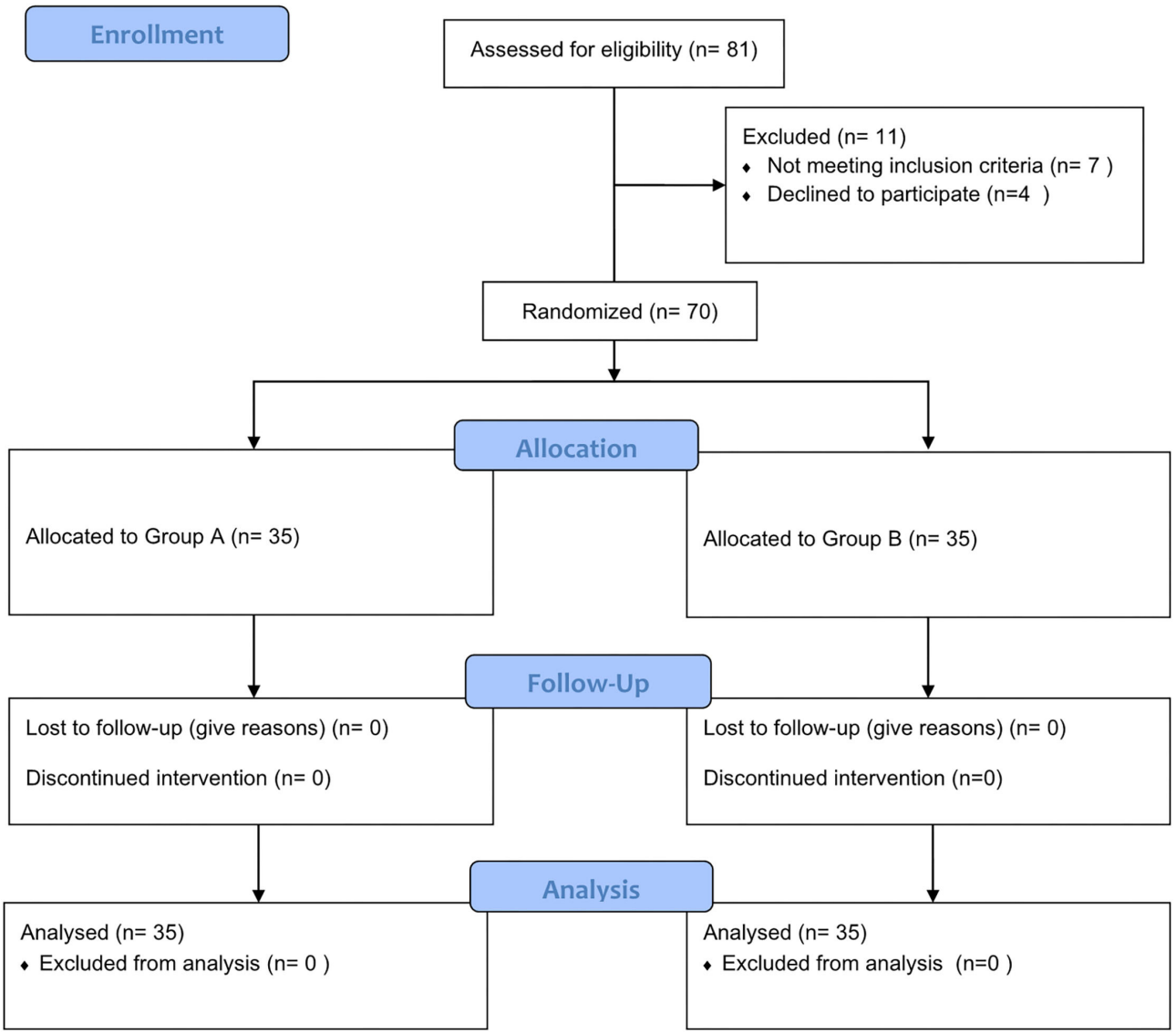


FIGURE 1 | Flowchart for the study recruitment.

at three time points (baseline, immediately after drug infusion, and immediately after intubation).

The primary outcomes included intubation times (from the begging of cricothyroid membrane puncture to the end of nasal tracheal intubation); intubation scores as assessed according to the vocal cord movement (1, open; 2, moving; 3, closing; 4, closed), coughing (1, none; 2, slight; 3, moderate; 4, severe), and limb movement (1, none; 2, slight; 3, moderate; 4, severe); patient tolerance as assessed using a five-point fiberoptic intubation comfort score (1, no reaction; 2, slight grimacing; 3, heavy grimacing; 4, verbal objection; 5, defensive movement of head or hands), and a three-point score assessed immediately after nasotracheal intubation (1, cooperative; 2, minimal resistance; 3, severe resistance); and first intubation attempt success rate.

Other parameters during intubation included the satisfaction of the operator regarding the intubation process (0, completely dissatisfied; 10, completely satisfied) and the occurrence of a hypoxic episode (SpO_2 < 92%). All the adverse events were recorded.

STATISTICAL ANALYSIS

We used the GraphPad Prism version 8.0 (GraphPad Software Inc., San Diego, California, USA) for the statistical analyses. Continuous variables are described as mean ± SD and were compared by using the paired *t*-test. The chi-squared test or the Fisher's exact test was used to compare categorical variables between the groups. Intubation conditions and tolerance score were analyzed using the independent samples Mann–Whitney *U* test. Blood pressure and HR at different time points were compared by using the two-way repeated measures analysis of variance. A $p < 0.05$ was regarded as statistically significant.

RESULTS

A total of 70 patients (33 males and 37 females) with anticipated difficult airways were enrolled in this study (**Figure 1**). The baseline data of the two groups showed no differences (**Table 1**).

The Ramsay Sedation Scale (RSS) score after the drug infusion showed no significant differences between both the groups. All the patients were successfully intubated with AFOI. There was no significant difference between both groups in the first intubation attempt success rate, but the first intubation attempt failed in two patients from Group A. The intubation time in Group B (200.4 ± 28.1) was lower than in Group A (244.1 ± 91.5) ($p = 0.003$). The intubation scores were better in Group B compared with Group A with regard to the vocal cord movement, cough, limb movement, and operator satisfaction ($p < 0.05$), but there was no significant difference in the patient tolerance score and comfort score of the patients after intubation (**Table 2**).

There was no significant difference in the hemodynamic change between both the groups at three time points: baseline, immediately after drug infusion, and immediately after intubation (**Figure 2**).

The incidence of adverse events (hypertension, hypotension, tachycardia, bradycardia, and hypoxia) was not significantly different between both the groups. Postanesthetic interview parameters: hoarseness, sore throat, and recall of intubation did not differ between both the groups, but patient satisfaction in Group B (9.4 ± 0.8) was better compared with that in Group A (7.5 ± 1.3) ($p = 0.0007$) (**Table 3**).

TABLE 1 | Demographic and clinical characteristics of study participants.

Characteristic	Group A ($n = 35$)	Group B ($n = 35$)	*P* value
Age (years)	57.6 ± 9.8	56.7 ± 11.9	0.734
Sex (male/female)	16/19	17/18	0.811
Weight (Kg)	66.2 ± 9.8	65.2 ± 10.1	0.657
ASA status (1/2/3)	16/16/3	12/20/3	0.602
Mallampatti grade (2/3/4)	6/19/10	8/20/7	0.841
Mouth opening (cm)	3.6 ± 0.6	3.8 ± 0.7	0.184
RSS score (1/2)	8/27	6/29	0.766

Data are presented as mean ± standard (SD) or number.
ASA, American Society of Anesthesiologists.

TABLE 2 | Airway management characteristics.

	Group A ($n = 35$)	Group B ($n = 35$)	*P* value
RSS after drug infusion (2/3)	9/26	12/23	0.603
First intubation attempt success rate (%)	94.286	100	0.151
Intubation time (sec)	244.1 ± 91.5	200.4 ± 28.1	0.003
Intubation scores			
Vocal cord movement 1/2/3/4	12/17/4/0	22/12/1/0	0.01
Cough 1/2/3/4	14/18/3/0	27/8/0/0	0.002
Limb movement 1/2/3/4	22/6/7/0	29/5/1/0	0.002
Patient tolerance 1/2/3/4	10/20/5/0	17/17/1/0	0.29
Patients' comfort score after intubation 1/2/3/4	24/10/1/0	30/4/1/0	0.267
Operator's satisfaction (1–10)	7 ± 0.5	8.9 ± 0.9	0.0001

Data are presented as mean ± standard (SD) or number.
RSS, Ramsay sedation score.

DISCUSSION

This study aimed to compare the two techniques of topical anesthesia that are both deemed effective during AFOI in patients with anticipated difficult airways. To the best of our knowledge, our study is the first to compare cricothyroid membrane puncture anesthesia and topical anesthesia in awake fiberoptic nasotracheal intubation.

In the existing literature on AFOI, several viewpoints exist. Some studies mention that cricothyroid membrane puncture anesthesia is a more effective method than topical anesthesia using the spray-as-you-go technique (18, 22), while others state that the spray-as-you-go technique is superior to the cricothyroid membrane puncture technique (23).

This study showed that both the cricothyroid membrane puncture anesthesia and topical anesthesia were effective for AFOI, but the cricothyroid membrane puncture anesthesia technique required less intubation time and led to better intubation scores, a higher success rate of first intubation,

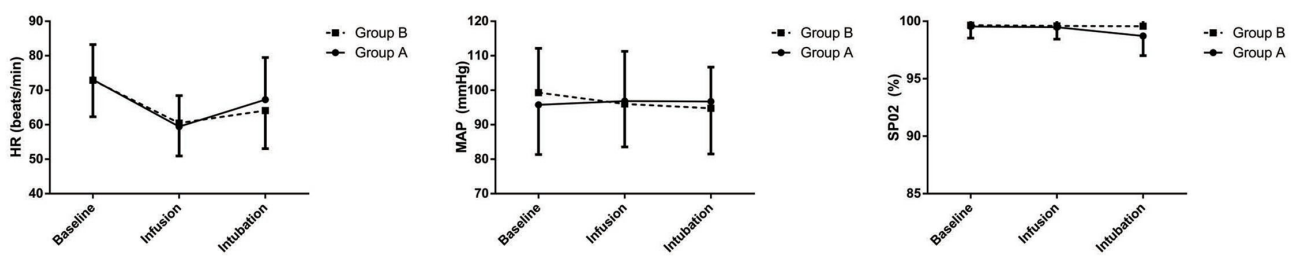


FIGURE 2 | Heart rate (HR), mean arterial blood pressure (MAP), and peripheral oxygen saturation (SpO_2) before premedication (baseline), immediately after drug infusion (infusion), and immediately after intubation (intubation). Group A: Topical anesthesia; Group B: Cricothyroid membrane puncture anesthesia. Points are expressed as mean ± SD.

TABLE 3 | Postoperative visit data and advent events.

	Group A ($n = 35$)	Group B ($n = 35$)	*P* value
Hoarseness	1	1	1
Sore throat	1	3	0.614
Recall of intubation	25	30	0.244
Patients' satisfaction	7.5 ± 1.3	9.4 ± 0.8	0.0007
Hypertension	0	0	1
hypotension	0	0	1
tachycardia	1	1	1
bradycardia	2	3	0.69
hypoxia	0	0	1

Data are presented as mean ± standard (SD) or number.

and a higher satisfaction score for both operator and patients during AFOI.

We chose the intubation time as the primary outcome of our study, because it is the clearest criterion to evaluate the efficiency of AFOI. Along with an adequate analgesia and sedation, we believe that a rapid AFOI procedure is important for ensuring safety and comfort of the patient. Our study found that the cricothyroid membrane puncture anesthesia technique was significantly faster than the topical anesthesia technique; this is consistent with the results of a previous study (24). Moreover, to ensure the double blindness of the experiment, the cricothyroid membrane puncture anesthesia group also used the spray-as-you-go technique for comparison with topical anesthesia; this increased the intubation time by 2 min and might have reduced patient comfort and satisfaction.

The secondary outcomes were the intubation scores during AFOI. Scoring systems similar to the one we used to evaluate intubation scores are described in the literature (6, 25, 26), but the results were different in those studies. Our study showed that the cricothyroid membrane puncture anesthesia technique was better than the topical anesthesia technique as it exhibited better intubation scores, higher success rate of first intubation, and higher operator satisfaction score. Although cricothyroid membrane puncture anesthesia was administered subglottically, it provided a better topical block. Despite the fact that the patient tolerance and patient comfort scores were better in the cricothyroid membrane puncture anesthesia group, this difference did not achieve statistical significance.

We knew from our experience and from the literature (27–29) that adequate analgesia and sedation are necessary for AFOI. Therefore, we administered sufentanil (0.2 μg/kg) and dexmedetomidine (1 μg/kg) over 10 min before AFOI to obtain sufficient analgesia and sedation. On following this procedure, all the patients had an RSS score > 1 after the drug infusion and no patient felt obvious pain in any of the AFOI procedures; this result is consistent with those of the previous studies (30–32).

Hemodynamic stability is a measure of stress response during AFOI. We did not find any significant difference between the two groups regarding this factor. This finding is consistent with the results of the previous studies, wherein AFOI performed by experienced operators had no influence on the hemodynamic stability of the patient (33, 34).

Dexmedetomidine has been reported to decrease noradrenaline release and centrally mediated sympathetic tone (35). However, it may cause side effects including hypotension, bradycardia, and hypoxia (36). In our study, five patients developed bradycardia after drug infusion that can be treated easily with atropine and two patients developed tachycardia after intubation. There was no significant difference between the two groups regarding the adverse effects of dexmedetomidine.

Postoperative visit data showed the majority of the patients had no hoarseness or soreness of the throat. The cricothyroid membrane puncture anesthesia group exhibited higher patient satisfaction and less recall of intubation than the topical anesthesia group. Therefore, we conclude that cricothyroid membrane puncture anesthesia technique, if performed by experienced anesthesiologists, provide a high level of comfort to the patient.

This study had some limitations. There were only 70 patients enrolled in our study, further larger sample studies are required to confirm our results. Another limitation was the lack of dose-effect study; the doses of sufentanil and dexmedetomidine used were based on data provided in the previous studies. Our drug infusion protocol is not suitable for the emergency operation, as it requires more than 10 min. Despite sufficient analgesia and sedation, cricothyroid membrane puncture is still an invasive operation.

CONCLUSION

Both the techniques of topical anesthesia are effective in AFOI, but cricothyroid membrane puncture anesthesia provided better intubation conditions and less intubation time with greater satisfaction of the patient and operator when compared with topical anesthesia.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Review Board of the Ethics Committee of Shanghai Tenth People's Hospital affiliated with Shanghai Tongji University School of Medicine (SHSY-IEC-4.0/19-81-01). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SW, CH, TZ, XZ, and CL: provision of study materials or patients, collection and assembly of data, and final

approval of manuscript. CL: administrative support. SW, CH, XZ, and CL: conception and design, data analysis and interpretation, and manuscript writing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant Number 81600921) located in No. 83 Shuangqing Road, Haidian District, Beijing, 100085, China and the Natural Science Foundation of Shanghai (Grant Number 20ZR1442900) located in No. 200, Renmin Avenue, Shanghai, 200003, China.

ACKNOWLEDGMENTS

We sincerely thank the Department of Anesthesiology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, 200072, China. The Department of Anesthesiology provided all the medical apparatus and instruments for this study.

REFERENCES

1. Shiga T, Wajima Z, Inoue T, Sakamoto A. Predicting difficult intubation in apparently normal patients: a meta-analysis of bedside screening test performance. *Anesthesiology.* (2005) 103:429–37. doi: 10.1097/00000542-200508000-00027
2. Schmitt H, Buchfelder M, Radespiel-Tröger M, Fahlbusch R. Difficult intubation in acromegalic patients: incidence and predictability. *Anesthesiology.* (2000) 93:110–4. doi: 10.1097/00000542-200007000-00020
3. Wang S, Hu C, Zhang T, Zhao X, Li C. Practice guidelines for management of the difficult airway: an updated report by the American Society of anesthesiologists task force on management of the difficult airway. *Anesthesiology*. (2003) 98:1269–77. doi: 10.1097/00000542-200305000-00032
4. Apfelbaum JL, Hagberg CA, Caplan RA. Practice guidelines for management of the difficult airway: an updated report by the American Society of anesthesiologists task force on management of the difficult airway. *Anesthesiology.* (2013) 118:251–70. doi: 10.1097/ALN.0b013e31827773b2
5. Dang BW, Zhang J. Safety and efficacy of argon plasma coagulation for resection of lipomas and hamartomas in large airways. *Asian Pacific J Cancer Prevent. APJCP.* (2011) 12:477–80. doi: 10.1097/01.cad.0000390767.85658.83
6. Li CW Li YD, Tian HT, Kong XG, Chen K. Dexmedetomidine-midazolam versus Sufentanil-midazolam for awake fiberoptic nasotracheal intubation: a randomized double-blind study. *Chin Med J.* (2015) 128:3143–8. doi: 10.4103/0366-6999.170260
7. Chu KS, Wang FY, Hsu HT, Lu IC, Wang HM, Tsai CJ. The effectiveness of dexmedetomidine infusion for sedating oral cancer patients undergoing awake fibreoptic nasal intubation. *Eur J Anaesthesiol.* (2010) 27:36–40. doi: 10.1097/EJA.0b013e32832e0d2b
8. Shen SL, Xie YH, Wang WY, Hu SF, Zhang YL. Comparison of dexmedetomidine and sufentanil for conscious sedation in patients undergoing awake fibreoptic nasotracheal intubation: a prospective, randomised and controlled clinical trial. *Clin Respir J.* (2014) 8:100–7. doi: 10.1111/crj.12045
9. Bergese SD, Patrick Bender S, McSweeney TD, Fernandez S, Dzwonczyk R, Sage K, et al. comparative study of dexmedetomidine with midazolam and midazolam alone for sedation during elective awake fiberoptic intubation. *J Clin Anesth.* (2010) 22:35–40. doi: 10.1016/j.jclinane.2009.02.016
10. Stamenkovic DM, Hassid M. Dexmedetomidine for fiberoptic intubation of a patient with severe mental retardation and atlantoaxial instability. *Acta Anaesthesiol Scand.* (2006) 50:1314–5. doi: 10.1111/j.1399-6576.2006.01157.x
11. Maroof M, Khan RM, Jain D, Ashraf M. Dexmedetomidine is a useful adjunct for awake intubation. *Canad J Anaesthesia.* (2005) 52:776–7. doi: 10.1007/BF03016576
12. Isaac PA, Barry JE, Vaughan RS, Rosen M, Newcombe RG, A. jet nebuliser for delivery of topical anesthesia to the respiratory tract. a comparison with cricothyroid puncture and direct spraying for fibreoptic bronchoscopy. *Anaesthesia.* (1990) 45:46–8. doi: 10.1111/j.1365-2044.1990.tb14504.x
13. Koerner IP, Brambrink AM. Fiberoptic techniques. Best practice and research. *Clinic Anaesthesiol.* (2005) 19:611–21. doi: 10.1016/j.bpa.2005.07.006
14. Andruszkiewicz P, Dec M, Kański A, Becler R. Awake fibreoptic intubation. *Anestezjologia intensywna terapia.* (2010) 42:218–21.
15. Andruszkiewicz P, Dec M, Kański A, Becler R. Fibreoptic intubation in awake patients. *Anestezjologia intensywna terapia. (*2010) 42(4):194-196.
16. Pirlich N, Lohse JA, Noppens RR. Topical airway anesthesia for awake-endoscopic intubation using the spray-as-you-go technique with high oxygen flow. *JoVE.* (2017) 13:119. doi: 10.3791/55116
17. Sidhu VS, Whitehead EM, Ainsworth QP, Smith M, Calder I, A. technique of awake fibreoptic intubation. experience in patients with cervical spine disease. *Anaesthesia.* (1993) 48:910–3. doi: 10.1111/j.1365-2044.1993.tb07429.x
18. Webb AR, Fernando SS, Dalton HR, Arrowsmith JE, Woodhead MA, Cummin AR. Local anaesthesia for fibreoptic bronchoscopy: transcricoid injection or the "spray as you go" technique? *Thorax.* (1990) 45:474–7. doi: 10.1136/thx.45.6.474
19. Peiris K, Frerk C. Awake intubation. *J Perioper Pract.* (2008) 18:96–104. doi: 10.1177/175045890801800302
20. Madan K, Mittal S, Gupta N, Biswal SK, Tiwari P, Hadda V, et al. The Cricothyroid versus spray-as-you-go method for topical anesthesia during flexible bronchoscopy: the crisp randomized clinical trial. *Respirat Int Rev Thorac Dis.* (2019) 98:440–46. doi: 10.1159/000501563
21. Chandra A, Banavaliker JN, Agarwal MK. Fibreoptic bronchoscopy without sedation: is transcricoid injection better than the "spray as you go" technique? *Indian J Anaesth.* (2011) 55:483–7. doi: 10.4103/0019-5049.89877
22. Graham DR, Hay JG, Clague J, Nisar M, Earis JE. Comparison of three different methods used to achieve local anesthesia for fiberoptic bronchoscopy. *Chest.* (1992) 102:704–7. doi: 10.1378/chest.102.3.704
23. Sethi N, Tarneja VK, Madhusudanan TP, Shouche S. Local anaesthesia for fiberoptic intubation: a comparison of three techniques. *Med J Armed Forces India.* (2005) 61:22–5. doi: 10.1016/S0377-1237(05)80112-1
24. Malcharek MJ, Bartz M, Rogos B, Günther L, Sablotzki A, Gille J, et al. Comparison of Enk Fibreoptic Atomizer with translaryngeal injection for topical anaesthesia for awake fibreoptic intubation in patients at risk of secondary cervical injury: a randomised controlled trial. *Eur J Anaesthesiol.* (2015) 32:615–23. doi: 10.1097/EJA.0000000000000285
25. Xu T, Li M, Ni C, Guo XY. Dexmedetomidine versus remifentanil for sedation during awake intubation using a Shikani optical stylet: a randomized, double-blinded, controlled trial. *BMC Anesthesiol.* (2016) 16:52. doi: 10.1186/s12871-016-0219-9
26. Hu R, Liu JX, Jiang H. Dexmedetomidine versus remifentanil sedation during awake fiberoptic nasotracheal intubation: a double-blinded randomized controlled trial. *J Anesth.* (2013) 27:211–7. doi: 10.1007/s00540-012-1499-y
27. Liu HH, Zhou T, Wei JQ, Ma WH. Comparison between remifentanil and dexmedetomidine for sedation during modified awake fiberoptic intubation. *Exp Ther Med.* (2015) 9:1259–64. doi: 10.3892/etm.2015.2288
28. Lallo A, Billard V, Bourgain JL, A. comparison of propofol and remifentanil target-controlled infusions to facilitate fiberoptic nasotracheal intubation. *Anesth Analg.* (2009) 108:852–7. doi: 10.1213/ane.0b013e318184eb31
29. Machata AM, Gonano C, Holzer A, Andel D, Spiss CK, Zimpfer M, et al. Awake nasotracheal fiberoptic intubation: patient comfort, intubating conditions, and hemodynamic stability during conscious sedation with remifentanil. *Anesth Analg.* (2003) 97:904–8. doi: 10.1213/01.ANE.0000074089.39416.F1

30. Wang JC, Yang TS, Chen PH, Huang WH, Hsiao SY, Wong CS, et al. Enhanced perioperative safety and comfort during airway-related surgeries and procedures with dexmedetomidine-a brief review and clinical practice experience. *Asian J Anesthesiol.* (2018) 56:56–63. doi: 10.6859/aja.201806_56(2).0003
31. Mirkheshti A, Memary E, Honar BN, Jalaeefar A, Sezari P. The efficacy of local dexmedetomidine during fiberoptic nasotracheal intubation: A randomized clinical trial. *J Anaesthesiol Clin Pharmacol.* (2017) 33:209–14. doi: 10.4103/joacp.JOACP_242_16
32. Wang HL, Tang SH, Wang XQ, Gong WH, Liu XM, Lei WF. Doxapram hastens the recovery following total intravenous anesthesia with dexmedetomidine, propofol and remifentanil. *Exp Ther Med.* (2015) 9:1518–22. doi: 10.3892/etm.2015.2249
33. Piepho T, Thierbach AR, Gobler SM, Maybauer MO, Werner C. Comparison of two different techniques of fibreoptic intubation. *Eur J Anaesthesiol.* (2009) 26:328–32. doi: 10.1097/EJA.0b013e32831ac4ce
34. Sutherland AD, Williams RT. Cardiovascular responses and lidocaine absorption in fiberoptic-assisted awake intubation. *Anesth Analg.* (1986) 65:389–91. doi: 10.1213/00000539-198604000-00016
35. Bloor BC, Ward DS, Belleville JP, Maze M. Effects of intravenous dexmedetomidine in humans. ii hemodynamic changes. *Anesthesiology.* (1992) 77:1134–42. doi: 10.1097/00000542-199212000-00014
36. Ebert TJ, Hall JE, Barney JA, Uhrich TD, Colinco MD. The effects of increasing plasma concentrations of dexmedetomidine in humans. *Anesthesiology.* (2000) 93:382–94. doi: 10.1097/00000542-200008000-00016

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 03 December 2021
doi: 10.3389/fmed.2021.775047

An Interpretable Early Dynamic Sequential Predictor for Sepsis-Induced Coagulopathy Progression in the Real-World Using Machine Learning

Ruixia Cui[1,2†], Wenbo Hua[3†], Kai Qu[1], Heran Yang[3], Yingmu Tong[1,2], Qinglin Li[1,2], Hai Wang[1,2], Yanfen Ma[4], Sinan Liu[1,2], Ting Lin[1,2], Jingyao Zhang[1,2,5], Jian Sun[3*] and Chang Liu[1,2,5*]*

[1] Department of Hepatobiliary Surgery, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [2] Department of SICU, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [3] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China, [4] Department of Clinical Laboratory, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China, [5] Biobank, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

Sepsis-associated coagulation dysfunction greatly increases the mortality of sepsis. Irregular clinical time-series data remains a major challenge for AI medical applications. To early detect and manage sepsis-induced coagulopathy (SIC) and sepsis-associated disseminated intravascular coagulation (DIC), we developed an interpretable real-time sequential warning model toward real-world irregular data. Eight machine learning models including novel algorithms were devised to detect SIC and sepsis-associated DIC $8n$ ($1 \leq n \leq 6$) hours prior to its onset. Models were developed on Xi'an Jiaotong University Medical College (XJTUMC) and verified on Beth Israel Deaconess Medical Center (BIDMC). A total of 12,154 SIC and 7,878 International Society on Thrombosis and Haemostasis (ISTH) overt-DIC labels were annotated according to the SIC and ISTH overt-DIC scoring systems in train set. The area under the receiver operating characteristic curve (AUROC) were used as model evaluation metrics. The eXtreme Gradient Boosting (XGBoost) model can predict SIC and sepsis-associated DIC events up to 48 h earlier with an AUROC of 0.929 and 0.910, respectively, and even reached 0.973 and 0.955 at 8 h earlier, achieving the highest performance to date. The novel ODE-RNN model achieved continuous prediction at arbitrary time points, and with an AUROC of 0.962 and 0.936 for SIC and DIC predicted 8 h earlier, respectively. In conclusion, our model can predict the sepsis-associated SIC and DIC onset up to 48 h in advance, which helps maximize the time window for early management by physicians.

Keywords: SIC, sepsis-associated DIC, irregular time-series data, early real-time prediction, machine learning

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Jesus Rico-Feijoo,
Hospital Universitario Río Hortega, Spain
Takeshi Wada,
Hokkaido University, Japan

***Correspondence:**
Jingyao Zhang
jingyaozhang@xjtu.edu.cn
Jian Sun
jiansun@xjtu.edu.cn
Chang Liu
liuchangfh@xjtu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *13 September 2021*
Accepted: *08 November 2021*
Published: *03 December 2021*

Citation:
Cui R, Hua W, Qu K, Yang H, Tong Y, Li Q, Wang H, Ma Y, Liu S, Lin T, Zhang J, Sun J and Liu C (2021) An Interpretable Early Dynamic Sequential Predictor for Sepsis-Induced Coagulopathy Progression in the Real-World Using Machine Learning. Front. Med. 8:775047. doi: 10.3389/fmed.2021.775047

INTRODUCTION

Sepsis is a lethal disease caused by a dysregulated host response in an infected state (1). Septic-induced organ dysfunction is a major cause of sepsis high mortality (2). Among these, coagulation dysfunction is a pervasive complication of sepsis, occurring in 50–70% of sepsis patients, while approximately 35% of patients proceed to disseminated intravascular coagulation (DIC) (3). Sepsis-induced coagulopathy (SIC) mortality reaches to 23.1% (4), while the mortality rate of sepsis-associated DIC is more than twice that of simple sepsis patients (5, 6). According to the SIC scoring system proposed by the DIC subcommittee of International Society on Thrombosis and Haemostasis (ISTH) in 2017, sepsis patients easily meet the SIC diagnostic criteria (7). Sepsis-associated DIC was diagnosed by the two-step sequential approach of SIC and ISTH overt-DIC criteria, which is a late-phase coagulation disorder that should be detected early (8). Currently, DIC diagnosis does not have a gold standard. Physicians diagnose DIC according to the primary disease, clinical manifestations, and laboratory tests. However, the clinical signs and symptoms of DIC appear slowly and the manifestations are complex and varied, resulting in a time lag in the clinical diagnosis of DIC, which places the patient in a treatment-refractory phase when sepsis-associated DIC is clinically determined (9). In addition, studies have shown that anticoagulation is ineffective in both sepsis and SIC patients, but effective in sepsis-induced DIC patients (10–12). Hence, early recognition of sepsis-associated DIC is more important than SIC, while there are currently no studies on sequential prediction of sepsis-associated DIC after SIC alerts. Therefore, it is imperative to establish a model for early sequential real-time prediction of SIC and sepsis-associated DIC.

The prevalence of electronic health records (EHRs) and the upsurge of artificial intelligence (AI) provide opportunities for clinical medical research (13). Studies have shown that machine learning-based models outperform traditional clinical scoring and human expert systems in the diagnosis, treatment, and prognosis prediction of clinical diseases (14, 15). However, current clinical prediction studies are mainly static and lack real-time prediction studies. Real-time prediction models dynamically predict the onset of disease within a sliding time-window by continuously updating clinical data. From the clinical dynamic treatment perspective, real-time predictive models would better fit the clinical applications (16). In addition, the variability of primary diseases, comorbidities and severity of conditions in ICU patients leads to sparse and irregular clinical data in terms of sampling time and dimensions (17). To accommodate irregular time series data, the existing standard models such as eXtreme Gradient Boosting (XGBoost) (18) and Recurrent Neural Network (RNN) (19) decompose time into continuous, non-overlapping uniform intervals, known as temporal discretization (20). This enables the standard models to act on fixed dimensional vectors (regular data). However, this approach lacks the continuity principle and can lead to undesirable results when applied to irregular medical time-series data (21). Altogether, it is necessary to develop a model that is specifically designed to handle sparse irregular time series data in the real clinical world to achieve real-time accurate predictions at arbitrary time points.

In summary, we aim to help physicians to identify patients at high risk of SIC and sepsis-associated DIC early, especially those who progress to DIC after SIC, as well as improve existing machine learning models to enable arbitrary time-point prediction on real-world irregular data. To achieve this, an interpretable early real-time sequential warning predictor will be developed that contributes to early personalized treatment and reasonable administration. The overview of the study design and model development was shown in **Figure 1**.

METHODS

Study Cohort and Design

This is a multicenter retrospective cohort study. Research data were obtained from two medical centers, the Xi'an Jiaotong University Medical College (XJTUMC) and Beth Israel Deaconess Medical Center (BIDMC). Structured query language (SQL) was used to obtain eligible patient data for the period from January 1, 2013 to December 1, 2018 in XJTUMC and from 2001 to 2012 in BIDMC, respectively. The XJTUMC data were obtained from the Center's Biobank and the BIDMC data were obtained from the Medical Information Mark for Intensive Care (MIMIC-III) database (22). The study was reviewed by the Ethics Committee of the First Affiliated Hospital of Xi'an Jiaotong University, and all data were deidentified.

The enrollment process was shown in **Figure 2**. The inclusion criteria were as follows: (1) sepsis was diagnosed within 24 h of admission based on Sepsis 3.0 criteria; (2) the age was not <18 years; and (3) the duration of hospitalization was not <3 days. The exclusion criteria were as follows: (1) patients with DIC onset within 24 h of admission; (2) patients were affected by hematologic tumors (leukemia, lymphoma, etc.); (3) patients suffering from cirrhosis, acute liver failure, with liver function up to Child C; (4) patients treated with radiotherapy or chemotherapy; (5) patients with admission diagnosis of combat trauma, traumatic coagulopathy; (6) patients with pregnancy or perinatal complications.

Data Preprocessing

We initially obtained a total of 174 laboratory features in XJTUMC and 259 laboratory features in BIDMC. Subsequently, we performed the process of merging identical variables (e.g., HGB with different units of g/L or g/dl, HGB in blood count and blood gas test, etc.), eliminating irrelevant variables (e.g., hepatitis antibody quantification, blood drug concentration, etc.), and counting the frequency of variable detection. We removed indictors that were completely missing and detected <1% of indicators at all-time points. In addition, we performed the unit conversions for the XJTUMC laboratory test variables, in order to maintain consistency with BIDMC. Ultimately, under the guidance of the laboratory physicians, we identified 99 features at XJTUMC and 72 features at BIDMC. Of these, the 72 features at BIDMC were common features for both medical centers. The

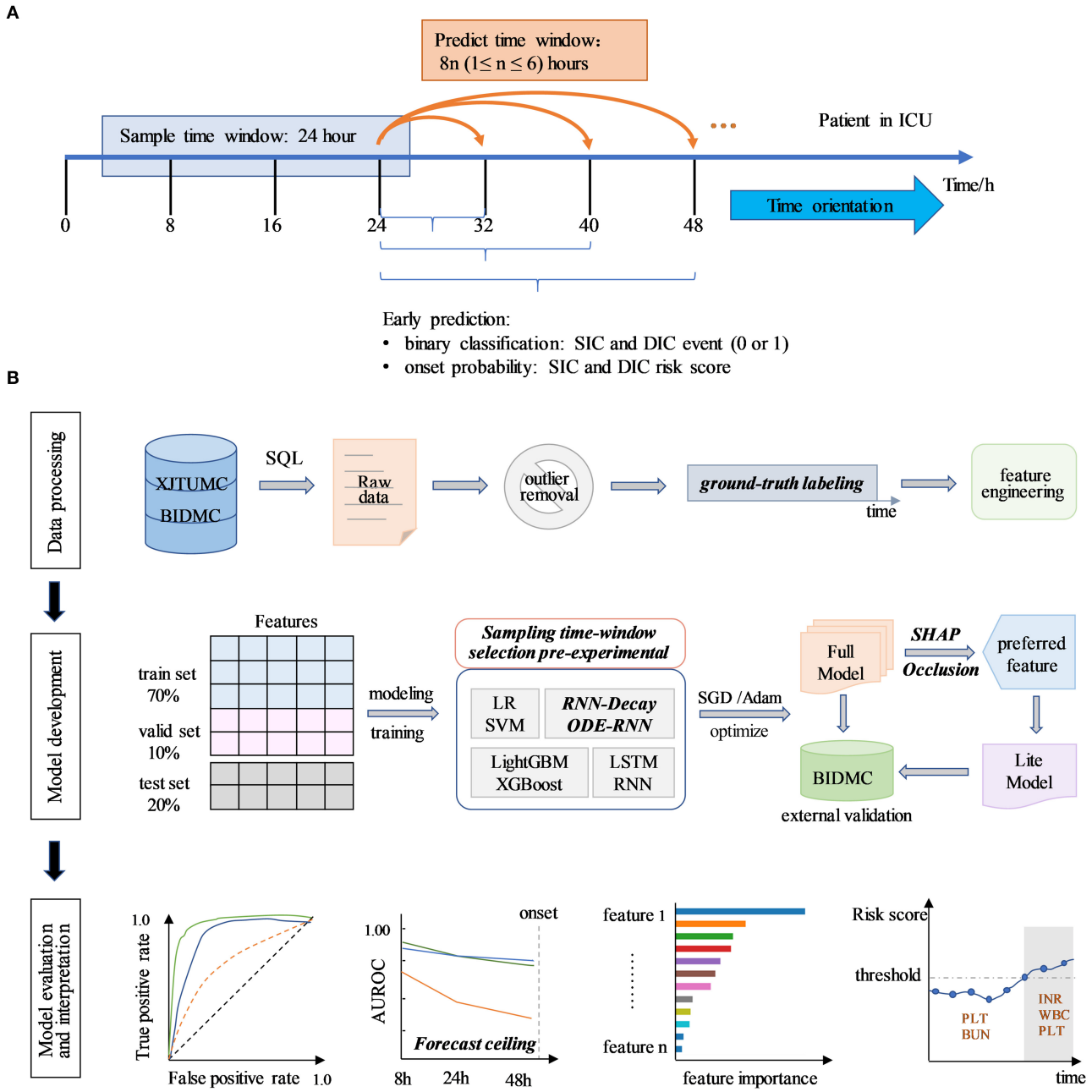


FIGURE 1 | Study design and model development overview. **(A)** Research application design profile. Study developed an early real-time prediction model to dynamically predict disease progression of sepsis patients after admission to ICU, and to output whether the patient will suffer from SIC and sepsis-associated DIC and the risk score for their onset in future periods. The model feeds into the information obtained from a continuous 24 h sampling time-window and outputs the disease status for the next 8*n* h in real-time. The x-axis indicates the time since admission. **(B)** Model development overview. The research process comprises to data processing, model development, model evaluation, and interpretation. The data processing process includes raw data extraction, outlier processing, coagulation status annotation, and feature engineering. The model development stage included feature set split, sampling time-window selection pre-experimental, model construction, and training. We built eight models, trained and selected the Full model with good performance on the internal validation set by SGD or Adam optimization. The Lite models were developed by selecting easily accessible and important features through SHAP and Occlusion analysis. Models were validated at BIDMC for external validation. The model evaluation and interpretation section include the assessment of model predictive performance, interpretation of model outputs, individualized disease trajectory prediction, etc.

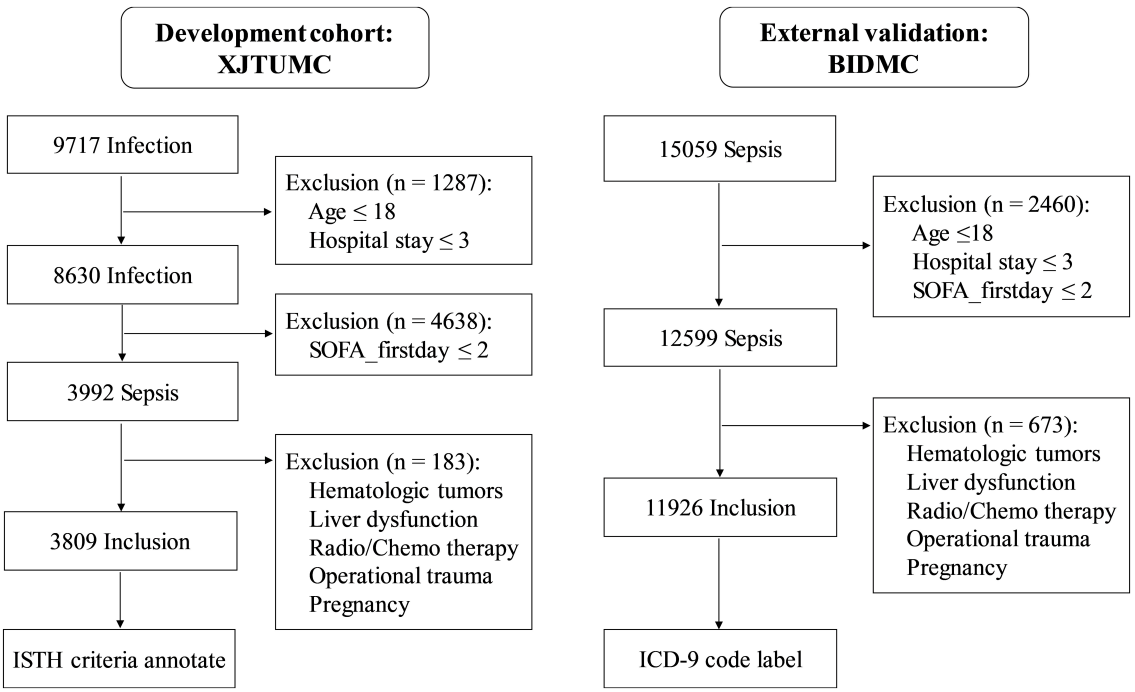


FIGURE 2 | The diagram of study patient enrollment. In XJTUMC, coagulation status annotation was based on SIC and ISTH-overt DIC criteria. In BIDMC, DIC were labeled based on ICD-9. ICD, International Classification of Diseases.

missing information for both medical centers were shown in the **Supplementary Figure 1** (**Supplementary File 1**). And the clinical reportable ranges for each identified variable was shown in **Supplementary File 2**. All variables were initialized by the min-max normalization algorithm.

Clinical sequence data are sparse and irregular, manifested by the presence of a large amount of missing data. Whereas, the pattern of missing data contains important information, such as the correlation between the frequency of a certain marker test and the severity of the disease. Therefore, in this study, we do not deal with missing data. Instead, we use the modeling of missing information to identify the role of missing patterns in prediction, which enhances the model prediction effectiveness. In brief, it is modeled by learning to characterize the missing and hidden information from the time-series data, which is then further introduced into the model network. The multivariate time series $X = \{x_1, x_2, \ldots, x_T\}$ is the observations at time T. $x_t \in R^D$represents the observations at time t for all variables. x_t contains D features $\left\{x_t^1, x_t^2, x_t^D\right\}$ and x_t^d represents the d-th feature variable of x_t. s_t denotes the timestamp of observation x_t. We assume that the timestamp of the first observation is 0 ($s_t = 0$), the time interval between different timestamps may be the same or different. Δ indicates adjacent time steps for each variable. To provide an efficient representation of the missing values, we introduce the mask vector $m_t^d = \{0, 1\}$ to represent the missing variables in x_t at time t. Some features are missing continuously over a period of time, and δ_t^d is defined to represent the time interval between the last observation and the current timestamp. To be more specific, we have:

$$m_t^d = \begin{cases} 1, & if\ x_t^d\ is\ observed \\ 0, & otherwise \end{cases} \tag{1}$$

$$\delta_t^d = \begin{cases} s_t - s_{t-1} + \delta_{t-1}^d & t > 1,\ m_{t-1}^d = 0 \\ s_t - s_{t-1} & t > 1,\ m_{t-1}^d = 1 \\ 0 & t = 1 \end{cases} \tag{2}$$

Thus, a dynamic feature of the input would be represented as $X_t = (x_t^d, m_t^d, \Delta, \delta_t^d)$. In the later section, the missing information will be introduced into the model for subsequent processing when improving the algorithm model.

Ground-Truth Labels Using ISTH

We defined all three disease states according to SIC and ISTH overt-DIC criteria. The details of the disease status annotation criteria can be found in **Supplementary Table 1** (**Supplementary File 1**). New data were not available for all 8 h time-window of the day because the patient's laboratory tests were irregular. If there are no updated data available for labeling in an 8 h time-window, then the forward interpolation method is used for labeling based on the labels before and after that time-window. More specific details were shown in **Supplementary Figure 2** (**Supplementary File 1**).

Continuous Models for Irregular Time Series

The study compared two methods for dealing with irregular time series. The first is the temporal discretization approach, in which the standard RNN model is a typical model. In the second approach, we tackle the irregular time-series problem by modeling the missing information. Based on the standard RNN model, we introduced a decay mechanism for modeling missing information by referencing Che et al.'s study (21), to eventually develop the RNN-Decay model. In addition, we developed the Ordinary Differential Equations-Recurrent Neural Networks (ODE-RNN) model by directly modeling the unequally spaced raw data referring to the work of Yulia et al. (23). The model architecture diagrams were shown in **Figure 3.** The model inputs are derived from the extraction of missing patterns of time series information from "Data Processing" described in the above section.

Model Development

To select the appropriate sampling time length, we chose 8, 24, and 48 h sampling time-window to perform pre-experiments. The results were presented in the **Supplementary Table 2** (**Supplementary File 1**). Finally, we set the sampling time-window to 24 h and the sliding time-step to 8 h. Then, we developed several state-of-the-art models that are widely used as follows: (1) Classic machine learning models: Logistic regression (LR) (24) and support vector machines (SVM) (25), which are the most commonly used algorithms in existing research; (2) Enhanced machine learning models: gradient boosting machine (LightGBM) (26) and XGBoost (27), which are widely regarded as the best algorithm for data prediction and are adopted by many competition winning models in the field of machine learning; (3) Classic deep learning models: RNN (21) and long short-term memory network (LSTM) (28), which are the most commonly chosen deep learning models in time-series data, which have shown excellent performance in several time series studies; (4) Improved deep learning models: RNN-Decay and ODE-RNN. The detailed method was described in **Supplementary File 1**. After evaluating the performance, we finally chose the highest-performing XGBoost model as our predictor.

Model Evaluation

We random divided the XJTUMC data into a train set (70%), a valid set (10%), and a test set (20%). The BIDMC data ware used as the external validation set to evaluate generalization ability of the model. Parameter optimization were performed by Stochastic Gradient Descent (SGD) or the Adam algorithm. The area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC) and the F1-score were used as model evaluation metrics. Test samples were resampled 1,000 times using bootstrapping to calculate 95% confidence intervals.

Model Interpretation

In this study, we interpreted the machine learning model results using the Shapley Additive exPlanations (SHAP) algorithm (29) and the deep learning model results using Occlusion analysis (30).

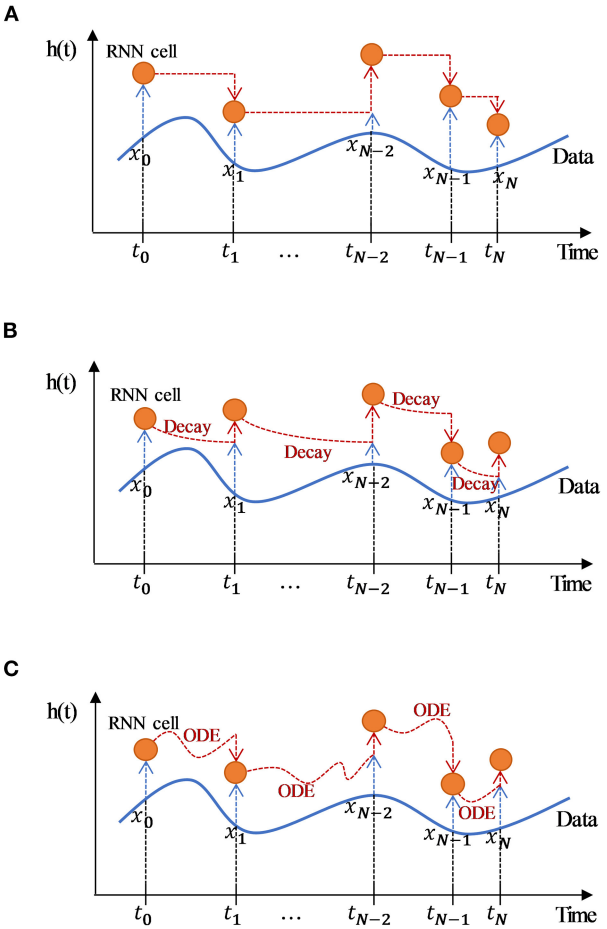


FIGURE 3 | Architecture diagram of different RNN-based models for processing irregular time-series data**. (A)** Diagram of the standard RNN model processing irregular time-series data. The standard RNN transmits the hidden layer information forward at a fixed time interval without processing the hidden layer information; **(B)** Diagram of RNN-Decay model processing irregular time-series data. Based on the RNN framework, RNN-decay adds temporal decay derived from time intervals to the latent space in order to indirectly process irregular series; however, the longer the time interval, the weaker the information conveyed. **(C)** Diagram of ODE-RNN model processing irregular time-series data. The ODE-RNN directly modeled the unequally spaced raw data, and propagated a latent variable along the time interval using the Neural ODE to solve the latent layer information that corresponds to the irregular timestamp; this allows the arbitrary time-point data to be processed and guarantees temporal continuity. The time series $T = \{t_0, t_1, \ldots, t_N\}$ is the observation timestamps, $X = \{x_0, x_1, \ldots, x_N\}$ represents the observations at time t for all variables, $h(t)$ is the latent space state corresponding to timestamp t. Orange circles indicate an RNN unit. Blue and black dashed lines with arrows indicate the shallow and latent features of the input RNN, and orange dashed lines with arrows indicate data from the previous RNN unit that has been processed in a different way and then fed into the next RNN unit.

Statistical Analysis

Baseline data are skewed and expressed as the median and interquartile range (IQR). Non-parametric tests were used for

statistical tests. *P*-values< 0.001 were considered statistically significant. Navicat Premium (12.1.22), Pytorch (1.7.0), and Python (3.7.6) with Numpy (1.18.5), Pandas (1.1.5), and Scikit-learn (0.23.2) formed the data-processing pipeline. All computational analyses were performed in the Computer Center of the School of Mathematics, Xi'an Jiaotong University.

RESULTS

Study Baseline

A total of 9,717 infection patients in XJTUMC and 15,059 sepsis patients in BIDMC were initially included when applying ICD-9 codes and sepsis 3.0 criteria. After applying the exclusion criteria, 3,809 and 11, 926 patients were left. We then annotated the coagulation status of 3,809 XJTUMC sepsis patients by SIC and ISTH overt-DIC criteria, a total of 12,154 SIC status labels (positive: 8,909; negative: 3,246); and 7,878 overt-DIC status labels (positive: 3,051; negative: 4,827); were available in XJTUMC. Also, we selected 296 patients (1,210 status labels) who developed DIC during hospitalization at the BIDMC center using ICD-9 codes. The baseline characteristics of included patients at XJTUMC and BIDMC were shown in **Table 1**. For both XJTUMC and BIDMC, the median age of sepsis patients is above 60 years and the predominant cause of sepsis was respiratory system-derived infections. Also, 652/1,415 (46.1%) of SIC patients developed to DIC, while 652/679 (96%) of DIC patients fulfilled the SIC diagnosis.

TABLE 1 | Baseline characteristics of included patients at XJTUMC and BIDMC.

	XJTUMC (n = 3,809)	BIDMC (n = 11,926)
Demographic		
Age (year), median [Q1, Q3]	63 [52,72]	69 [57,80]
Male, *n* (%)	2,388 (62.7)	6,481 (54.3)
Severity status at admission		
SOFA score, median [Q1, Q3]	3 [3, 4]	4 [5, 8]
Infection sources in sepsis, *n* (%)		
Respiratory system	1,702 (44.7)	4,957 (41.6)
Gastrointestinal system	1,435 (37.7)	3,194 (26.8)
Urinary system	5 (0.1)	382 (3.2)
Cardiac bloodstream system	24 (0.6)	728 (6.1)
Oncology cachexia related	567 (14.9)	1,077 (9.0)
Other	76 (2.0)	1,588 (13.3)
Outcome, median [Q1, Q3]		
Hospital stay (day)	7 [10, 15]	13 [7, 23]
Coagulation status		
SIC onset, *n* (%)	1,415 (37.1)[a]	Unknown
DIC onset, *n* (%)	679 (17.8)[a]	296 (2.5)[b]

For infection sources in sepsis, respiratory system infections mainly cover the lung, trachea, bronchus, and chest related diseases; gastrointestinal system infections typically involve the esophageal, gastric, bowel, liver, spleen, and abdominal related disorders; urinary system infections include renal, ureteral, bladder, and urethra inflammatory diseases; cardiac bloodstream system comprise the cardiac, vascular, vascular catheter, and systemic infection related diseases; oncology cachexia is a variety of malignant diseases and cachectic manifestations; other infections consist of various inflammatory states of unknown etiology and brain diseases.

[a] *Diagnosis based on SIC and ISTH-overt DIC criteria for coagulation status annotation.*

[b] *DIC diagnosis based on International Classification of Diseases-9th edition (ICD9) codes (ICD9 diagnosis code for DIC is 2866).*

Full Model Performance

The predictive performance of eight different models for early DIC onset were shown in **Figure 4**. **Figures 4A,B** showed that XGBoost produced the best prediction performance (AUROC: 0.955; AUPRC: 0.939) and was validated in LightGBM, followed by ODE-RNN (AUROC: 0.936; AUPRC: 0.902). However, ODE-RNN ensures continuity of model prediction, which is more suitable for clinical applications than XGBoost. **Figure 4C** revealed that the performance of the XGBoost on the BIDMC external validation set has decreased (AUROC: 0.865). **Figure 4D** provided prediction performance of the model at different prediction time-window, revealing that our model could detect the event as early as 48 h before the ground-truth. **Figure 5** illustrated the prediction performance for SIC, showing similar results to DIC. **Figures 5A,B** showed that XGBoost provided the best prediction performance (AUROC: 0.973; AUPRC: 0.979). **Figure 5C** revealed that the performance of the XGBoost on the BIDMC external validation set has decreased (AUROC: 0.973). **Figure 5D** revealed that XGBoost could detect the event as early as 48 h before the ground-truth with AUROC reached 0.929. The detailed predictive performance of models at different prediction time-window was shown in **Table 2**. The model prediction performance decreases steadily as the prediction time-window extends. The XGBoost and ODE-RNN still maintain good prediction performance for SIC and DIC in the 48 h ahead of prediction time-window. Furthermore, we examined the early warning performance of the models on SIC and DIC with different alert thresholds at the 8 h prediction time-window, the results were shown in **Supplementary Tables 3**, **4** (**Supplementary File 1**). That allows clinicians to select different thresholds according to the characteristics of the different stages of disease development and treatment needs.

Model Interpretation

To understand the contribution of the features to the model predictions, we interpreted the XGBoost predictions using Shapley values, which were presented in **Figure 6**. **Figure 6A** showed the top 20 features that contribute most to the model output; **Figure 6B** showed the impact of the top 20 features on all samples in the model. Shaply analysis identified that the most valuable features for DIC prediction were platelet (PLT), D-dimer, International normalized ratio (INR), plateletcrit (PCT), fibrinogen (FIB), fibrin degradation products (FDP). We further developed dependency plots to capture the non-linear correlation between a single significant feature and the predicted risk. As an example, **Figure 6C** showed that when PLT was below 80, the shape-value was significantly increased with higher predicted risk. **Figure 6D** showed the interaction between PLT and PCT, where when PCT is low, the corresponding PLT feature value is low, SHAP takes a high value and the model output risk increases. In addition, the SHAP force plot provides insight into the output risk and decision factors for specific samples. In the case of **Figure 6E**, the model predicts the sample at a high risk of a DIC event based on PT, INR, and PLT.

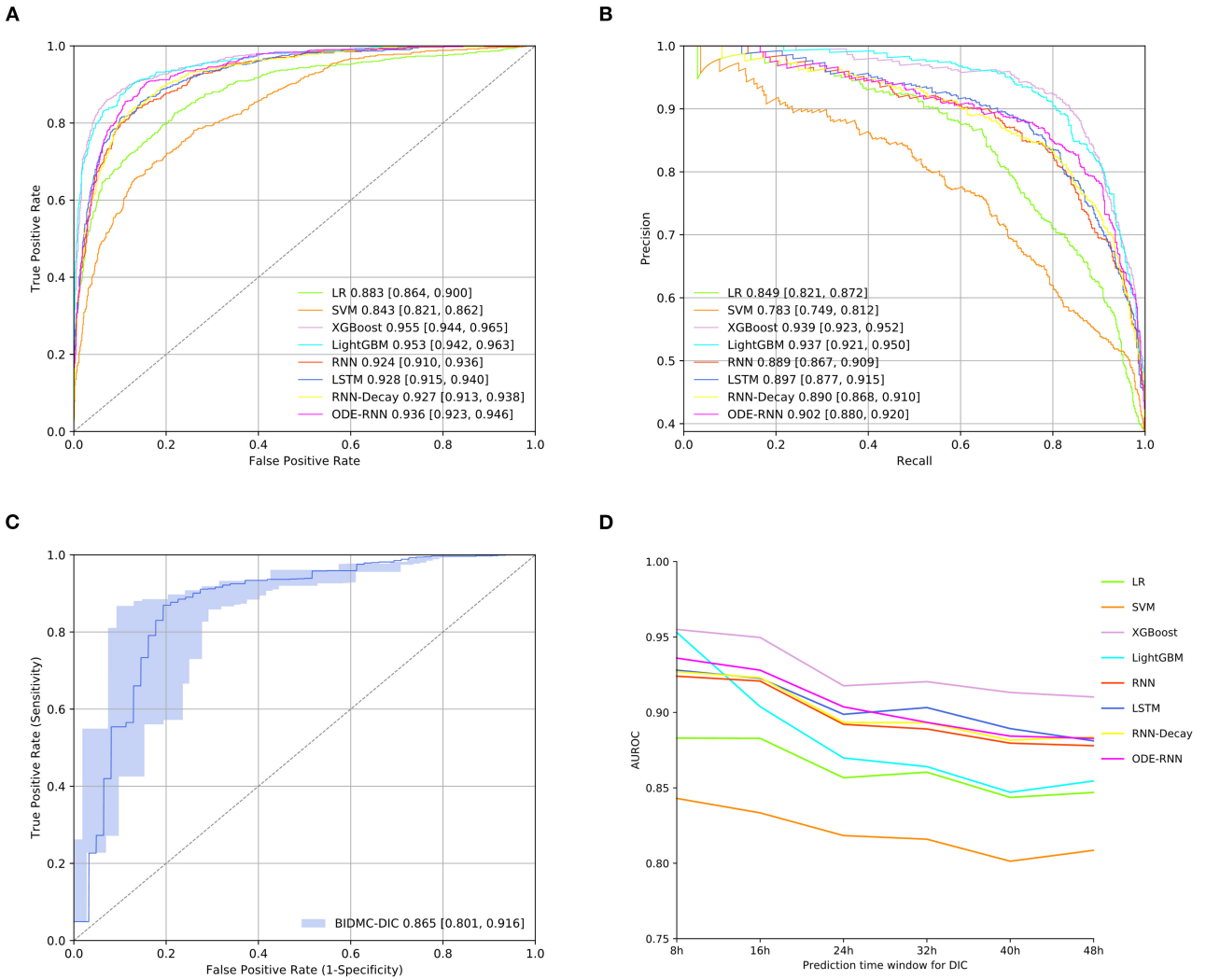


FIGURE 4 | Full model performance evaluation. **(A)** Receiver operating characteristic curves of models for DIC prediction on the internal test set. **(B)** Precision-recall curve of candidate models for DIC prediction on the internal test set. **(C)** Receiver operating characteristic curves on the external test set. **(D)** The predictive performance of models at 8, 24, and 48 h ahead of sepsis-associated DIC onset. Precision was defined as the fraction of correctly alerted 8 h before the ground-truth event. Recall was defined as the proportion of any alarms 8 h prior to the ground-truth event. The lighter-colored intervals above and below the ROC and PR curves are 95% confidence intervals.

In addition, we interpreted the deep learning model ODE-RNN using occlusion analysis, as shown in **Supplementary File 3**. Occlusion analysis showed that a global absence of a single feature has small impact on the model output results for the ODE-RNN model. We also performed univariate and multivariate analyses of baseline data on the first day for different coagulation status groups, as detailed in **Supplementary Tables 5**, **6** (**Supplementary File 1**), suggesting the reliability of the machine learning approach.

Figure 7 showed an example of real-time sequential prediction using our model on one patient. At each time point after the patient admission over 24 h, the model provides a real-time risk and uncertainty assessment of the future SIC and sepsis-associated DIC onset. This showed that the model could detect SIC and sepsis-associated DIC 48 h early, which is important for clinicians to take precautionary approaches ahead of the event.

Lite Model Development

To enhance the transferability and reduce the data requirements of the model, we selected the ten most influential features based on SHAP values, Occlusion analysis, and clinical practicability (The Lite model feature details were shown in **Supplementary File 2**). Based on the selected features, we constructed the Lite model. **Figure 8** indicated that the XGBoost-based Lite model achieves the best performance of predicting DIC 8 h in advance, but the model performance is slightly lower than that of the Full model (AUROC: 0.916 vs. 0.955).

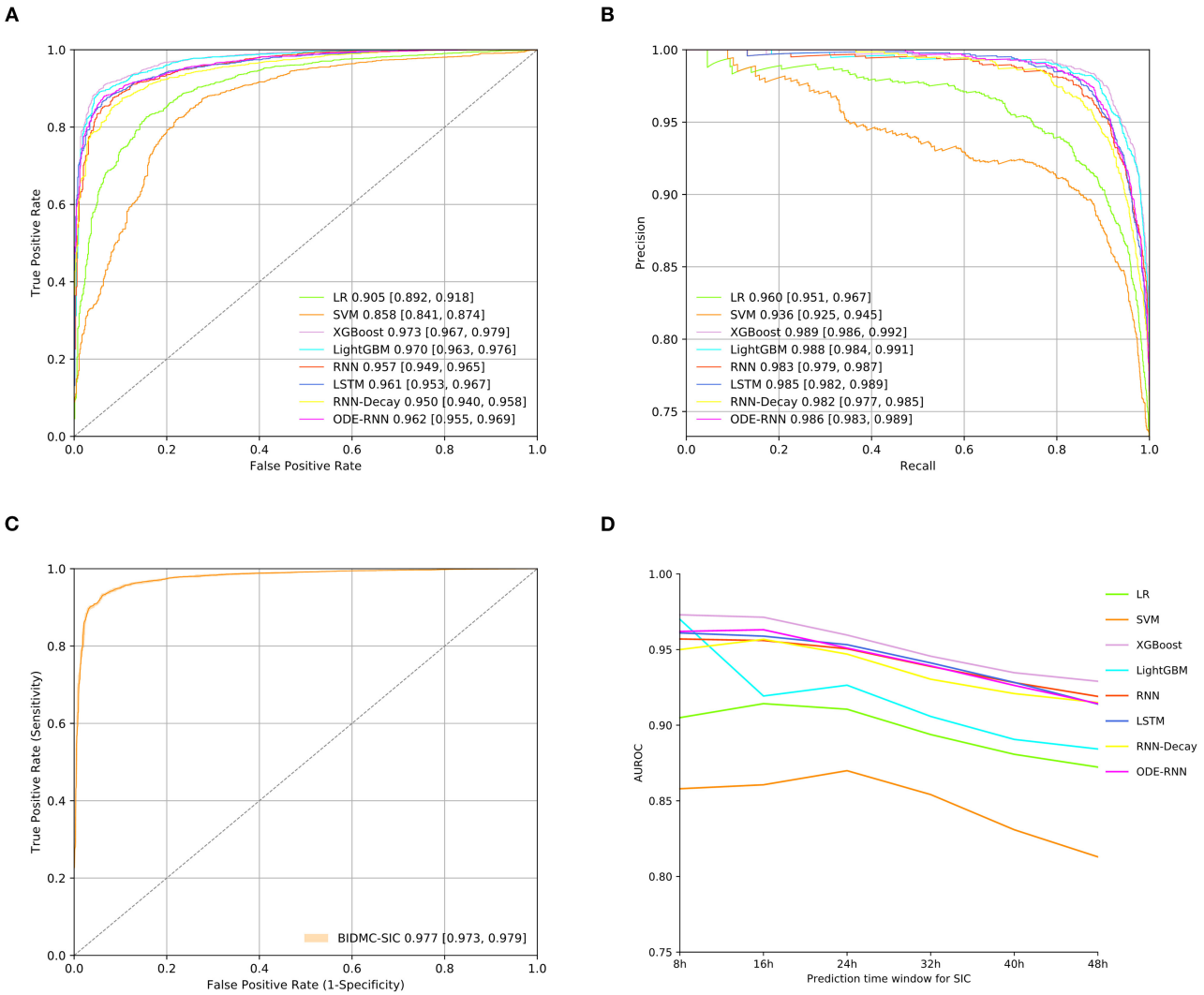


FIGURE 5 | Full model performance evaluation for SIC prediction. **(A)** Receiver operating characteristic curves of models for SIC prediction on the internal test set. **(B)** Precision-recall curve of candidate models for SIC prediction on the internal test set. **(C)** Receiver operating characteristic curves on the external test set. **(D)** The predictive performance of models at 8, 24, and 48 h ahead of SIC onset.

DISCUSSION

In this study, we developed two models enabling real-time sequential monitoring of SIC and DIC disease progression in sepsis. The model could identify high-risk patients 48 h before the clinical diagnosis of SIC and sepsis-associated DIC, achieving the state-of-the-art retrospective performance. The full XGBoost model currently achieved the highest prediction performance (SIC: 0.973; DIC: 0.955). The Lite XGBoost model also achieved pleasing prediction performance (DIC: 0.916). On the BIDMC test set, the model performance decreased slightly (DIC: 0.865). In the 8 h prediction time-window with a threshold of 0.7, the XGBoost model was able to predict 89.5% of SIC and 83.1% of sepsis-associated DIC events correctly. Meanwhile, our study has introduced a methodological improvement. Specifically, our study provided the following contributions: (1) We developed and validated the first model for earlier sequential dynamic monitoring of sepsis-induced coagulation disease progression; (2) We processed irregular time-series data for the first time in dynamic prediction research using ODE method, achieving predictions at arbitrary time points; (3) We provided the visual interpretation for deep learning models and machine learning models, respectively, which improved the recognition of physicians toward complex models.

Sepsis-induced SIC, particularly sepsis-associated DIC, is a major cause of increased mortality in sepsis. However, clinical DIC diagnosis relies on FIB and D-Dimer whose laboratory testing frequency is <10% (31), resulting in a lag in DIC diagnosis. Therefore, it is meaningful to use a

TABLE 2 | Model performance at different prediction time-window for SIC and DIC prediction.

Model-AUROC	SIC			DIC		
Prediction time-window	**8 h**	**24 h**	**48 h**	**8 h**	**24 h**	**48 h**
LR	0.905	0.911	0.872	0.883	0.857	0.847
SVM	0.858	0.870	0.813	0.843	0.818	0.809
XGBoost	0.973	0.960	0.929	0.955	0.918	0.910
LightGBM	0.970	0.926	0.884	0.953	0.870	0.855
RNN	0.957	0.951	0.919	0.924	0.892	0.878
LSTM	0.961	0.953	0.914	0.928	0.899	0.881
RNN-decay	0.950	0.947	0.915	0.927	0.893	0.884
ODE-RNN	0.962	0.951	0.914	0.936	0.904	0.883

SIC, sepsis-induced coagulopathy; DIC, disseminated intravascular coagulation; LR, logistic regression; SVM, support vector machines; LightGBM, light gradient boosting machine; XGBoost, eXtreme gradient boosting; RNN, recurrent neural network; LSTM, long short-term memory network; RNN-decay, recurrent neural networks-decay; ODE-RNN, ordinary differential equations-recurrent neural networks.

full spectrum of laboratory tests for the early detection of coagulation disorders. To our knowledge, there is only one study using machine learning to predict the progression of sepsis-induced DIC that was published in 2020 (32). Hasegawa et al. performed three classical machine learning methods to predict the progression of sepsis-induced coagulation disorders. In that study, sepsis was defined based on Systemic Inflammatory Response Syndrome (SIRS) criteria rather than sepsis 3.0 criteria. In addition, the study used the static data and the accuracy of the model to predict the progression of coagulopathy was only 59.8–67.0% (32). This is far poorer than our model prediction performance, which suggests that dynamic data monitoring is more consistent with clinical application than static models. In our study, high-risk patients were identified up to 48 h earlier, which suggests that comprehensive use of laboratory tests could detect coagulation disorders earlier. The Lite model also achieves satisfactory predictive performance.

The irregularity of clinical time series data was reflected in the tables as a large number of missing. Previous studies deal with large amounts of missing data by removing missing variables or using statistical interpolation methods, but such methods are not applicable to time series data (33). Neural ODE is a continuous dynamics theory that can explore the dynamic interactions between key features in the timeline of event onset and development (34). Our results demonstrate the superiority of ODE in the dynamics of disease. In a recent review, Alber et al. showed that it remains a challenge to apply ODE in medical continuous monitoring studies with incomplete baseline data and low-sampling data (35). Our study introduced ODE into RNNs and achieved better performance than RNN and LSTM, showing that ODE-RNNs are more appropriate for sparse irregular data than standard deep learning models. However, our results showed that the performance of the ODE-RNN was lower than that of the gradient boosted tree model (XGBoost and LightGBM). We consider that clinical laboratory diagnoses usually use hierarchical stratification for diagnosis, which fits better with the splitting structure of the tree model and gives the tree model a natural advantage. In addition, Qin et al. showed that neural models are not good at efficient feature transformation and scaling, while the tree-based model has an advantage in this respect (36). However, the flexibility and variety of tasks that can be achieved with deep learning are not available with traditional methods when dealing with complex tabular problems. Furthermore, occlusion analysis showed that a single feature masking does not affect the model significantly, indicating that ODE-RNN has better perturbation resistance. That is, the overall absence of a particular examination does not have a large impact on the ODE-RNN model, suggesting that the ODE-RNN model may have better robustness. Finally, the arbitrary time point continuous prediction achieved by ODE-RNN is also not possible with the gradient boosted tree model, where this arbitrary time point continuity is significant for clinical applications.

Our study offers the following potential benefits: Firstly, it is essential for ICU clinicians and nurses to identify patients who truly need intensive attention and personalized preventive medication. Our predictor can reduce the alarm frequency in patients without a high risk of sepsis-associated DIC occurrence after SIC. Secondly, our ODE-RNN model provides a reference for model selection of real-world irregular time series processing. This will facilitate subsequent studies to build robust models that better match the characteristics of clinical data. Thirdly, our model can be used to identify sepsis patients in different states of coagulation impairment, which could be useful for future randomized controlled clinical studies and further assist physicians to evaluate the time window of anticoagulation therapy appropriately. Finally, our interpretable model provides a visual interactive operating system prototype for early warning systems in ICU and will facilitate clinical deployment of predictive models.

However, some limitations also exist in our study. First of all, our model was developed in a single center, which reduces the effectiveness and may require retraining when the model is migrated to other centers. Furthermore, because we failed to obtain bedside real-time vital sign monitoring data, our model did not incorporate these parameters which may diminish the efficacy of the model. Finally, our study is retrospective and further prospective clinical studies need to be validated.

CONCLUSION

Our early dynamic sequential predictor enables identification of sepsis patients at high risk of SIC and DIC up to 48 h earlier, achieving the highest performance to date. Our study showed that the ODE-RNN model achieves better performance than the standard RNN model. Our study contributes to early personalized management, and also improves the currently available algorithms.

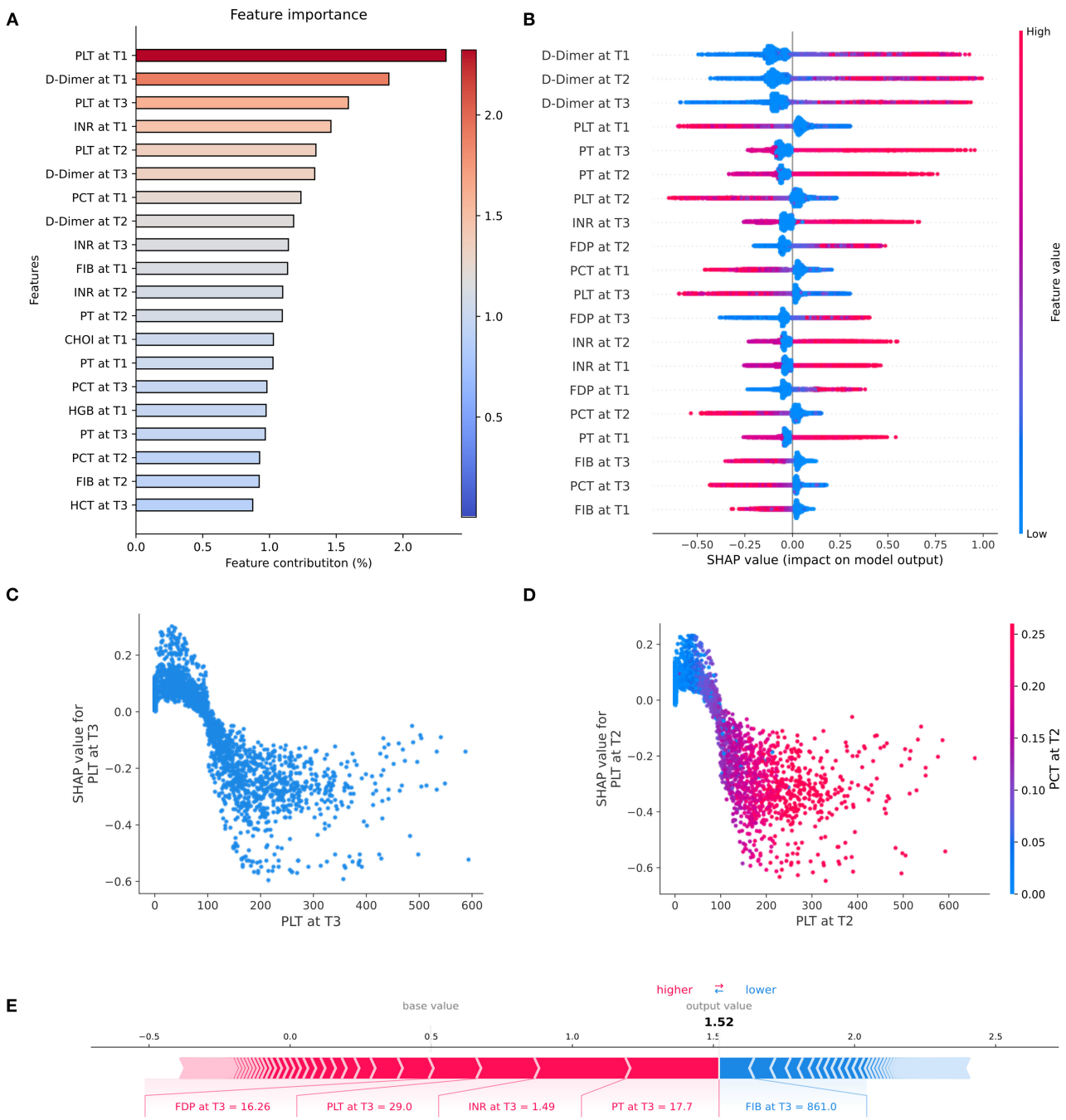


FIGURE 6 | SHAP-based interpretation of model. **(A)** Overall feature importance of the top 20 features. **(B)** Beeswarm plot showing the impact of the top 20 features across all samples in the model. Beeswarm combines feature importance and feature effect, ranking the features by the sum of the SHAP over all samples (y-axis). In the plot, one row represents one feature and each dot represents the feature SHapley value for one sample, gray dots represent missing values, colors represent feature values (red high, blue low). Long tails indicate that features are extremely important for a particular patient. The x-axis represents the impact on the model output, with positive values pushing risk higher and negative values driving risk lower. **(C)** SHAP dependence plot showing predicted risk vs. feature value for PLT. The x-axis is the range of eigenvalues of PLT features and the y-axis is the shape-value of PLT features. It showed that when PLT was below 80, the shape-value was significantly increased with higher predicted risk. **(D)** SHAP interaction dependency plot describes the relationship between two interaction features and predicted risk. It showed the interaction between PLT and PCT, where when PCT is low, the corresponding PLT feature value is low, SHAP takes a high value and the model output risk increases. **(E)** SHAP force plot shows an interpretable example of single-sample feature prediction results. The base value is the predicted average of the all sample, and the output value is the predicted value of the current sample. Red indicates that the feature increases the risk, blue indicates that the feature decreases the risk. Longer feature bars indicate that the feature is contributing more.

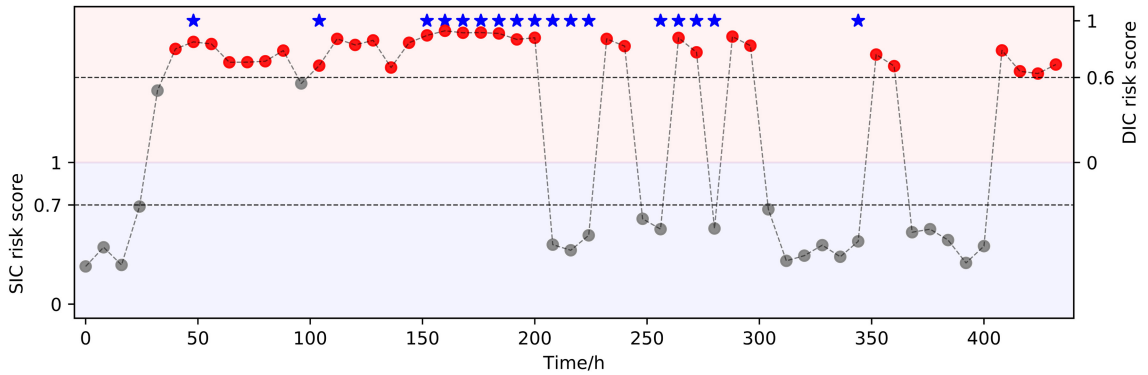


FIGURE 7 | An interpretable real-time risk forecasting example. A 69-year-old male patient was hospitalized with the diagnosis of "Acute biliary pancreatitis, Obstructive jaundice." The patient was diagnosed with SIC on the third day and sepsis-associated DIC on the fifth day of admission by the clinician. Our model predicted the disease progression earlier than the clinician. The predictions revealed that the most of sepsis-associated DIC events could be accurately predicted up to 8 h earlier. However, there were some sepsis-associated DIC events were still not predicted, mainly in the persistence of DIC onset, which might be related to clinical medicine interventions. The bottom of the graph is the SIC risk score and the upper is the DIC risk score. Red circles indicate sepsis-associated DIC events evaluated by the model. Red stars indicate true future predictions (meeting the ISTH overt-DIC criteria).

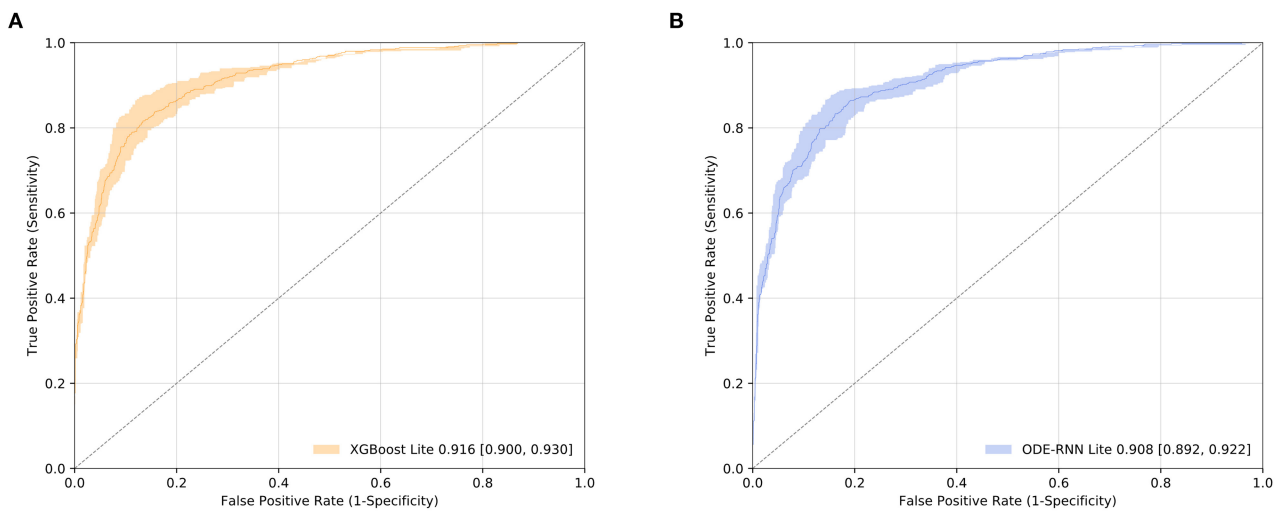


FIGURE 8 | Lite model performance evaluation. **(A)** Receiver operating characteristic curves of XGBoost for DIC. **(B)** Receiver operating characteristic curves of ODE-RNN for DIC.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by XJTU1AF2020LSL-003. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

RC designed the experiments, provided and preprocessed the clinical data, developed the supervised learning pipelines, and drafted the manuscript. WH designed the experiments, developed the machine learning pipelines, constructed the deep learning pipelines, and drafted the manuscript. KQ preprocessed the data and carried out the multivariate regression analysis. HY annotated the disease status labels and developed the deep learning pipelines. YT, QL, and HW carried out the data filtering and univariate analysis. YM, SL, and TL defined the clinically reportable range and performed multivariate regression analysis. JZ, JS, and CL revised the manuscript, conceived, and directed the project. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Critical Clinical Research Project of the First Affiliated Hospital of Xi'an Jiaotong University (No. XJTU1AF-CRF-2020-003) and the Joint Project of Universities in Shaanxi Province - Key Project (No. 2021GXLH-Z-099).

ACKNOWLEDGMENTS

We gratefully acknowledge the data support from the Biobank of First Affiliated Hospital of Xi'an Jiaotong University, and the MIMIC project from Massachusetts Institute of Technology and the Beth Israel Deaconess Medical Center. We also thank the helpful discussions with Yan Yang, Ph.D., and Dong Yang, Ph.D., School of Mathematics and Statistics, Xi'an Jiaotong University. We thank the technician Hui Liu who performed the data quality control, Biobank, the First Affiliated Hospital of Xi'an Jiaotong University. Finally, thanks to all reviewers for their time and suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.775047/full#supplementary-material

REFERENCES

1. Coopersmith CM, De Backer D, Deutschman CS, Ferrer R, Lat I, Machado FR, et al. Surviving sepsis campaign: research priorities for sepsis and septic shock. *Intensive Care Med.* (2018) 44:1400–26. doi: 10.1007/s00134-018-5175-z
2. Prescott HC, Angus DC. Enhancing recovery from sepsis: a review. *JAMA*. (2018) 319: 62–75. doi: 10.1001/jama.2017.17687
3. Levi M, van der Poll T. Coagulation and sepsis. *Thromb Res.* (2017) 149:38–44. doi: 10.1016/j.thromres.2016.11.007
4. Iba T, Umemura Y, Watanabe E, Wada T, Hayashida K, Kushimoto S, et al. Diagnosis of sepsis-induced disseminated intravascular coagulation and coagulopathy. *Acute Med Surg.* (2019) 6:223–32. doi: 10.1002/ams2.411
5. Voves C, Wuillemin WA, Zeerleder S. International Society on Thrombosis and Haemostasis score for overt disseminated intravascular coagulation predicts organ dysfunction and fatality in sepsis patients. *Blood Coagul Fibrin.* (2006) 17:445–51. doi: 10.1097/01.mbc.0000240916.63521.2e
6. Lyons PG, Micek ST, Hampton N, Kollef MH. Sepsis-associated coagulopathy severity predicts hospital mortality. *Crit Care Med.* (2018) 46:736–42. doi: 10.1097/CCM.0000000000002997
7. Iba T, Nisio MD, Levy JH, Kitamura N, Thachil J. New criteria for sepsis-induced coagulopathy (SIC) following the revised sepsis definition: a retrospective analysis of a nationwide survey. *BMJ Open.* (2017) 7:e017046. doi: 10.1136/bmjopen-2017-017046
8. Iba T, Arakawa M, Di Nisio M, Gando S, Anan H, Sato K, et al. Newly proposed sepsis-induced coagulopathy precedes international society on thrombosis and haemostasis overt-disseminated intravascular coagulation and predicts high mortality. *J Intensive Care Med.* (2020) 35:643–9. doi: 10.1177/0885066618773679
9. Gando S, Levi M, Toh CH. Disseminated intravascular coagulation. *Nat Rev Dis Primers.* (2016) 2:16037. doi: 10.1038/nrdp.2016.37
10. Umemura Y, Yamakawa K, Ogura H, Yuhara H, Fujimi S. Efficacy and safety of anticoagulant therapy in three specific populations with sepsis: a meta-analysis of randomized controlled trials. *J Thromb Haemost.* (2016) 14:518–30. doi: 10.1111/jth.13230
11. Umemura Y, Yamakawa K. Optimal patient selection for anticoagulant therapy in sepsis: an evidence-based proposal from Japan. *J Thromb Haemost.* (2018) 16:462–4. doi: 10.1111/jth.13946
12. Iba T, Levy JH, Warkentin TE, Thachil J, van der Poll T, Levi M. Diagnosis and management of sepsis-induced coagulopathy and disseminated intravascular coagulation. *J Thromb Haemost.* (2019) 17:1989–94. doi: 10.1111/jth.14578
13. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* (2019) 20:e262–73. doi: 10.1016/S1470-2045(19)30149-4
14. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol.* (2017) 2:230–43. doi: 10.1136/svn-2017-000101
15. Harerimana G, Kim JW, Yoo H, Jang B. Deep learning for electronic health records analytics. *IEEE Access.* (2019) 7:101245–59. doi: 10.1109/access.2019.2928363
16. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* (2020) 46:383–400. doi: 10.1007/s00134-019-05872-y
17. Holzinger A, Jurisica I. Knowledge discovery and data mining in biomedical informatics: the future is in integrative, interactive machine learning solutions. In: Holzinger A, Jurisica I, editors. *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics. Lecture Notes in Computer Science*, Vol. 8401. Berlin; Heidelberg: Springer (2014). p. 1–18. doi: 10.1007/978-3-662-43968-5_1
18. Chen TQ, Guestrin C, Assoc Comp M. XGBoost: a scalable tree boosting system. In: *Kdd'16: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA (2016). p. 785–94. doi: 10.1145/2939672.2939785
19. Wang J, Zhang L, Guo Q, Yi Z. Recurrent neural networks with auxiliary memory units. *IEEE Trans Neural Netw Learn Syst.* (2017) 29:1652–61. doi: 10.1109/TNNLS.2017.2677968
20. Futoma J, Hariharan S, Heller K. Learning to detect sepsis with a multitask gaussian process RNN classifier. In: Doina P, Yee Whye T, editors. *Proceedings of the 34th International Conference on Machine Learning. Vol. 70. Proceedings of Machine Learning Research: PMLR*. Sydney, NSW (2017). p. 1174–82.
21. Che Z, Purushotham S, Cho K, Sontag D, Liu Y. Recurrent neural networks for multivariate time series with missing values. *Sci Rep.* (2018) 8:6085. doi: 10.1038/s41598-018-24271-9
22. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data.* (2016) 3:160035. doi: 10.1038/sdata.2016.35
23. Rubanova Y, Chen RTQ, Duvenaud D. Latent ODEs for irregularly-sampled time series. In: *Advances in Neural Information Processing Systems 32 (Nips 2019). Vol. 32*. Vancouver, BC (2019).
24. Christensen R. Linear and log-linear models. *J Am Stat Assoc.* (2000) 95:1290–3. doi: 10.1080/01621459.2000.10474332
25. Noble WS. What is a support vector machine? *Nat Biotechnol.* (2006) 24:1565–7. doi: 10.1038/nbt1206-1565
26. Ke GL, Meng Q, Finley T, Wang TF, Chen W, Ma WD, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R. *Advances in Neural Information Processing Systems 30, Vol. 30*. Long Beach, CA (2017).
27. Sagi O, Rokach L. Approximating XGBoost with an interpretable decision tree. *Inf Sci.* (2021) 572:522–42. doi: 10.1016/j.ins.2021.05.055
28. Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* (2019) 31:1235–70. doi: 10.1162/neco_a_01199
29. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems 30 (Nips 2017)*. Vol. 30. Long Beach, CA (2017). p. 4768–777.
30. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision*. Zurich: Springer (2014). p. 818–33. doi: 10.1007/978-3-319-10590-1_53

31. Favaloro EJ, Negrini D. Machine learning and coagulation testing: the next big thing in hemostasis investigations? *Clin Chem Lab Med.* (2021) 59:1177–9. doi: 10.1515/cclm-2021-0216
32. Hasegawa D, Yamakawa K, Nishida K, Okada N, Murao S, Nishida O. Comparative analysis of three machine-learning techniques and conventional techniques for predicting sepsis-induced coagulopathy progression. *J Clin Med.* (2020) 9:2113. doi: 10.3390/jcm9072113
33. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. *IEEE J Biomed Health Inform.* (2017) 22:1589–604. doi: 10.1109/JBHI.2017.2767063
34. Chen RTQ, Rubanova Y, Bettencourt J, Duvenaud D. Neural ordinary differential equations. In: *Advances in Neural Information Processing Systems 31 (Nips 2018). Vol. 31.* Montréal, QC (2018).
35. Alber M, Buganza Tepole A, Cannon WR De S, Dura-Bernal S, Garikipati K, et al. Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. *NPJ Digit Med.* (2019) 2:115. doi: 10.1038/s41746-019-0193-y
36. Qin Z, Yan L, Zhuang H, Tay Y, Pasumarthi RK, Wang X, et al. Are neural rankers still outperformed by gradient boosted decision trees? In: *International Conference on Learning Representations (ICLR 2021).* Vienna (2021).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fmed.2021.810195

Development and Validation of a Simplified Prehospital Triage Model Using Neural Network to Predict Mortality in Trauma Patients: The Ability to Follow Commands, Age, Pulse Rate, Systolic Blood Pressure and Peripheral Oxygen Saturation (CAPSO) Model

Yun Li[1,2†], Lu Wang[1,2†], Yuyan Liu[1,2†], Yan Zhao[2], Yong Fan[3], Mengmeng Yang[2], Rui Yuan[2], Feihu Zhou[2], Zhengbo Zhang[3] and Hongjun Kang[2*]*

[1] Medical School of Chinese PLA, Beijing, China, [2] Department of Critical Care Medicine, The First Medical Centre, Chinese PLA General Hospital, Beijing, China, [3] Center for Artificial Intelligence in Medicine, Chinese PLA General Hospital, Beijing, China

Objective: Most trauma scoring systems with high accuracy are difficult to use quickly in field triage, especially in the case of mass casualty events. We aimed to develop a machine learning model for trauma mortality prediction using variables easy to obtain in the prehospital setting.

Methods: This was a retrospective prognostic study using the National Trauma Data Bank (NTDB). Data from 2013 to 2016 were used for model training and internal testing, and data from 2017 were used for validation. A neural network model (NN-CAPSO) was developed using the ability to follow commands (whether GCS-motor was <6), age, pulse rate, systolic blood pressure (SBP) and peripheral oxygen saturation, and a new score (the CAPSO score) was developed based on logistic regression. To achieve further simplification, a neural network model with the SBP variable removed (NN-CAPO) was also developed. The discrimination ability of different models and scores was compared based on the area under the receiver operating characteristic curve (AUROC). Furthermore, a reclassification table with three defined risk groups was used to compare NN-CAPSO and other models or scores.

Results: The NN-CAPSO had an AUROC of 0.911(95% confidence interval 0.909 to 0.913) in the validation set, which was higher than the other trauma scores available for prehospital settings (all $p < 0.001$). The NN-CAPO and CAPSO score both reached the AUROC of 0.904 (95% confidence interval 0.902 to 0.906), and were no worse than other prehospital trauma scores. Compared with the NN-CAPO, CAPSO score, and the other trauma scores in reclassification tables, NN-CAPSO was found to more accurately classify patients to the right risk groups.

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Matej Strnad,
Maribor University Medical Centre, Slovenia
Farzad Rahmani,
Tabriz University of Medical Sceinces, Iran

***Correspondence:**
Hongjun Kang
doctorkang301@163.com
Zhengbo Zhang
zhengbozhang301@gmail.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *06 November 2021*
Accepted: *19 November 2021*
Published: *10 December 2021*

Citation:
Li Y, Wang L, Liu Y, Zhao Y, Fan Y, Yang M, Yuan R, Zhou F, Zhang Z and Kang H (2021) Development and Validation of a Simplified Prehospital Triage Model Using Neural Network to Predict Mortality in Trauma Patients: The Ability to Follow Commands, Age, Pulse Rate, Systolic Blood Pressure and Peripheral Oxygen Saturation (CAPSO) Model. Front. Med. 8:810195. doi: 10.3389/fmed.2021.810195

Conclusions: The newly developed CAPSO system simplifies the method of consciousness assessment and has the potential to accurately predict trauma patient mortality in the prehospital setting.

Keywords: trauma, in-hospital mortality, prehospital, triage, scoring system, machine learning

INTRODUCTION

Trauma remains one of the leading causes of death and disability worldwide (1). Patients with severe trauma often benefit from receiving treatment at a higher level of care (2). Therefore, it is important to identify patients with severe trauma in the prehospital setting to avoid delayed or inadequate treatment, especially after a mass casualty incident (MCI). However, in many cases, the prehospital phase of triage is time-constrained and aids to diagnosis are limited, and even when many ambulance personnel do not have the relevant specialization, the number of personnel is severely insufficient compared to the large number of casualties (3). Thus, investigating how to quickly and accurately determine the severity of injuries using the most accessible assessment methods is needed.

To date, many severity assessment methods applicable to the early stages of trauma have been proposed and validated, including scoring systems or predictive models, most of which were constructed based on logistic regression analysis. The Revised Trauma Score (RTS), which was first proposed in 1989, used respiratory rate (RR), systolic blood pressure (SBP), and Glasgow Coma Scale (GCS) to calculate the probability of survival and is still widely used today (4). The Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure (MGAP) score developed in 2010 used four variables to assess the severity of trauma and performed better than the RTS (5). In contrast, the Glasgow Coma Scale, Age, Systolic Blood Pressure (GAP) score was proposed in 2011, which was referenced for the establishment of the MGAP, performed no less well than the MGAP with the mechanism of trauma removed (6). The NTS (New Trauma Score) score used peripheral oxygen saturation (SpO_2) instead of respiratory rate in the RTS and improved the prediction of death in trauma patients (7). The Trauma Rating Index in Age, Glasgow Coma Scale, Respiratory rate and Systolic blood pressure (TRIAGES) score used a generalized additive model to delineate the interval of variables and had better performance than the GAP score with the addition of the respiratory rate variable (8). Composed by the mechanism of trauma RTS, Injury Severity Score (ISS) (9) and age, the Trauma and Injury Severity Score (TRISS) was able to predict trauma mortality accurately (10). Although TRISS can hardly be used in prehospital settings because of the complex assessment of ISS, it is often taken as a benchmark for comparison with other trauma scores. However, in mass casualty incidents, it is also difficult to have sufficient time and manpower to monitor and assess all vital signs and complete GCS scores of casualties. Without the use of assistive electronic devices, calculating scores at the scene also has the disadvantage of being time-consuming and error-prone. In addition, the reliability of complete GCS score is dependent on relevant training and education (11), and it is often difficult for nonprofessional personnel involved in triage to accurately assess the GCS (12).

Therefore, it is necessary to explore the optimization of the input variables that need to be evaluated prehospital, for example, by considering the simplification of the consciousness assessment method (13) or by eliminating the systolic blood pressure variable, which is relatively difficult to measure (14). Alternatively, scores can be calculated quickly and accurately with the help of electronic devices, or for better prediction, sophisticated machine learning models can be embedded in them. Machine learning models can often better handle complex non-linear interactions between variables and improve the accuracy of results by optimizing the error between predicted and observed results (15). Other studies have shown that using only the motor component of the GCS is a simple and valid assessment tool, and even determining whether a patient has the ability to follow commands (assessing whether the GCS-motor is <6) has been shown to be a potential alternative to the GCS in the prehospital phase (16). However, there are no valid machine learning models or scores using this approach developed for predicting the mortality of trauma patients in the prehospital setting. The purpose of this study is to investigate the development of a machine learning model and a new easy-to-use trauma score for prehospital trauma mortality prediction. This will be achieved by using the binary assessment of GCS-motor (GCS-m) score <6 and other accessible vital signs, of which the predictive performance is not inferior to the RTS, MGAP, GAP, and TRIAGES scores.

METHODS

Study Design and Setting

Data were obtained from the National Trauma Data Bank (NTDB), the largest trauma database in the United States, which was assembled by the American College of Surgeons (17). Reporting of this study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) Guideline (18). Permission to use these data was obtained from NTDB.

Selection of Participants

This study used data from 2013 to 2017 in the NTDB, totaling 4,112,308 cases. The type of trauma was limited to blunt and penetrating. Cases without emergency medical service (EMS) data were excluded. To improve the quality of the included data, cases with more than three missing variables in the seven variables of SBP, HR, RR, GCS eye-opening response, GCS speech score, GCS motor score, and total GCS score were excluded.

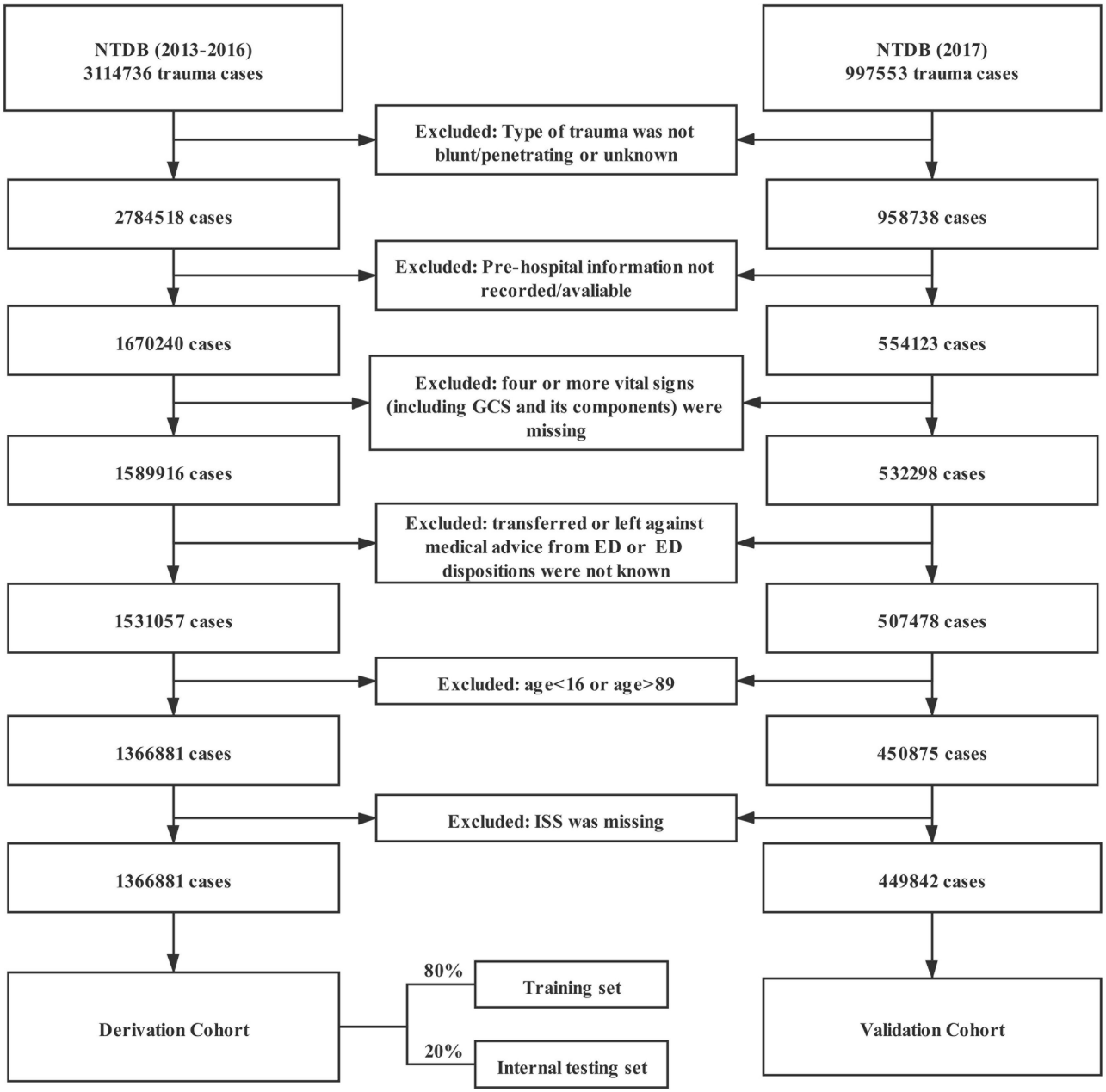


FIGURE 1 | Study participant selection procedure. *NTDB*, National Trauma Data Bank; *GCS*, Glasgow Coma Scale; *ED*, emergency department; and *ISS*, Injury Severity Score.

In addition, patients who were transferred from the emergency department (ED) to other hospitals, refused treatment in the ED, or had unknown outcomes in the ED were excluded. The age range of the patients was limited to 16 to 89 years (**Figure 1**).

Measurements and Outcome

Cases from 2013 to 2016 were used as the derivation cohort, and cases from 2017 were used as the validation cohort. Eighty percent of the derivation cohort was randomly assigned to the training set, and the remaining 20% was used as the internal testing set. Predictor variables for the study included age and vital signs that were first recorded in the field. Whether the GCS motor was <6, i.e., whether the patient had the ability to follow commands, was a simplified assessment of consciousness used in this study as an alternative to the GCS. The outcome variable was in-hospital death from any cause. The missing values in the derivation or validation cohort were imputed using multivariate imputation by chained equations (MICE) (19). In addition to the study variables, vital signs recorded in the emergency department (ED), type of trauma, injury severity score (ISS), length of

hospital stay, duration of mechanical ventilation, and length of intensive care unit (ICU) stay were also used for imputation. Due to the overlarge amount of data, only one imputed dataset was used for model development and validation.

Analysis

In the training set, the neural network algorithm and logistic regression analysis were used to develop mortality prediction models in trauma patients. The neural network consists of an input layer, hidden layers, and an output layer, where the neurons in each layer are first activated by neurons in the previous layer, then transformed by a non-linear function in the current layer, and eventually input to the next layer (20). This non-linear characteristic makes it efficient at learning complex relationships of input variables. Three models were developed using three combinations of variables based on the neural network respectively. The first combination of variables was GCS, age, pulse rate, SBP, and peripheral oxygen saturation, referred to as "GAPSO." The second combination of variables replaced GCS with a simpler binary assessment of GCS-Motor (GCS-m) score <6 (i.e., the ability to follow commands), referred to as "CAPSO." The third combination removed SBP from the second combination to further investigate the effect of removing blood pressure on the model's performance, referred to as "CAPO." The neural network models in this study contain two hidden layers with 256 and 128 neurons. They were optimized using the Adam optimizer, and overfitting was prevented by setting the dropout layer and early stopping.

Logistic regression analysis was performed, and then a score was developed using the second variable combinations, i.e., CAPSO. Considering the non-linear relationship between the variables and the outcome, the results of the multivariate generalized additive model were used to delineate the range of all predictors. Simple integers were assigned to the intervals according to the coefficients of the logistic regression, referring to the development of TRIAGES (8). Detailed methods for delineating variable intervals and assigning integer values are provided in the **Supplementary Material**. To be compared with the new models, the trauma scores previously developed were calibrated to the population in this study by fitting a logistic regression model to predict mortality for each score in the training set. Receiver operating characteristic (ROC) curves were plotted to compare the classification performance of each model as well as each score. The area under the receiver operating characteristic curves (AUROCs) were compared between different models using Delong's test (21). The agreement between the predicted probabilities of models or scores and observed frequencies of in-hospital mortality of trauma patients was assessed using probability calibration curves. To compare the sensitivity, specificity, and accuracy of the models and scores, it is necessary to select the threshold of mortality, whereby the prediction samples were classified into positive and negative samples. In this study, with reference to a previous study (5), the threshold with a sensitivity of at least 95% was set for comparison. Finally, to compare the differences between models when further classifying trauma patients, the trauma mortality predicted by each model and score was divided into three intervals as described in previous studies (5, 6): trauma patients at low (<5%), intermediate, and high (>50%) risk of death. The Shapley additive explanation (SHAP) plots (22) for the CAPSO model based on neural network were drawn. All statistical analyses were performed using Python (version 3.7.8) and R (version 4.0.2); neural network models were based on TensorFlow 2.1.0; $p < 0.05$ was considered statistically significant.

TABLE 1 | Baseline characteristics of trauma patients.

Variables	Derivation cohort (n = 1,366,881)	Validation cohort (n = 449,842)
Age, years [range]	52.0 [31.0, 70.0]	53.0 [32.0, 71.0]
Male, *n* (%)	843,371 (61.7)	273,196 (60.7)
Race, *n* (%)		
American Indian	11,196 (0.8)	3,492 (0.8)
Asian	25,923 (1.9)	9,285 (2.1)
Black or African American	204,017 (14.9)	68,516 (15.2)
Native Hawaiian or Other Pacific Islander	3,422 (0.3)	1,203 (0.3)
White	990,206 (72.4)	322,870 (71.8)
Other	132,117 (9.7)	44,476 (9.9)
Type of trauma, *n* (%)		
Blunt	1,230,013 (90.0)	404,176 (89.8)
Penetrating	136,868 (10.0)	45,666 (10.2)
First recorded vital signs measured at the scene of injury		
Systolic blood pressure, mmHg [range]	137.0 [120.0, 154.0]	138.0 [120.0, 156.0]
Pulse rate, beats/min [range]	89.0 [77.0, 102.0]	88.0 [76.0, 102.0]
Respiratory rate, rate/min [range]	18.0 [16.0, 20.0]	18.0 [16.0, 20.0]
Oxygen saturation, % [range]	98.0 [96.0, 99.0]	98.0 [95.0, 99.0]
Glasgow Coma Scale [range]	15.0 [14.0, 15.0]	15.0 [14.0, 15.0]
Injury Severity Score, [range]	9.0 [4.0, 13.0]	9.0 [4.0, 13.0]
Outcomes		
Length of stay in hospital, days [range]	4.0 [2.0, 7.0]	4.0 [2.0, 7.0]
ICU admission, *n* (%)	437,882 (32.0)	130,167 (28.9)
Mechanical ventilation, *n* (%)	209,471 (15.3)	55,315 (12.3)
Death, *n* (%)	65,770 (4.8)	22,208 (4.9)

Medians with 25th–75th interquartile ranges are shown for continuous variables, and counts with percentages are shown for categorical variables.

RESULTS

Characteristics of Study Subjects

Based on the inclusion and exclusion criteria, a total of 1,816,723 cases were included in the study, with 1,366,881 cases in the derivation cohort and 449,842 cases in the validation cohort (**Figure 1**). The main characteristics of the trauma patients after

TABLE 2 | Predictors at presentation associated with in-hospital death used to develop CAPSO in the derivation dataset.

Predictors	Beta [95% CI]	*P* value	Integerized score point
Intercept	−4.93 [−4.95, −4.90]	<0.001	
Age, years			
16–49	Reference		0
50–64	0.42 [0.39, 0.45]	<0.001	1
65–74	0.92 [0.88,0.95]	<0.001	2
75+	1.27 [1.24, 1.30]	<0.001	3
Glasgow Coma Scale-Motor			
<6	2.39 [2.37, 2.41]	<0.001	5
6	Reference		0
Systolic blood pressure, mmHg			
0–49	2.25 [2.19, 2.31]	<0.001	4
50–89	1.00 [0.87, 1.04]	<0.001	2
90–109	0.52 [0.49, 0.55]	<0.001	1
110–199	Reference		0
200+	0.54 [0.49, 0.59]	<0.001	1
Pulse rate, beats/min			
0–49	1.26 [1.21, 1.32]	<0.001	3
50–59	0.63 [0.58, 0.68]	<0.001	1
60–119	Reference		0
120–189	0.60 [0.57, 0.63]	<0.001	1
190+	1.39 [1.11, 1.66]	<0.001	3
Oxygen saturation, %			
0–79	1.50 [1.46, 1.54]	<0.001	3
80–89	0.95 [0.94, 0.98]	<0.001	2
90–94	0.41 [0.37, 0.44]	<0.001	1
95–100	Reference		0

95% CI, 95% confidence interval; mmHg, millimeter of mercury.

imputation of missing values are shown in **Table 1**. The overall median age of all cases was 52 years, and the interquartile range (IQR) was 31 to 70 years. A total of 61.5% of patients were male, and the overall mortality rate was 4.8%. The baseline characteristics before imputation of missing values are shown in **Supplementary Table 1**.

Development of Mortality Prediction Models

Based on the neural network algorithm, three models were developed using three sets of variable combinations respectively, including neural network-based GAPSO (NN-GAPSO), neural network-based CAPSO (NN-CAPSO), and neural network-based CAPO (NN-CAPO). The continuous variables in the predictor combination CAPSO were classified into categorical variables based on the generalized additive model. After analysis through logistic regression, the new score, CAPSO (the Ability to Follow Commands, Age, Pulse Rate, Systolic Blood Pressure, and peripheral Oxygen saturation), was defined after assigning integer values to the variables according to the coefficients of the regression equation. The CAPSO scores ranged from a maximum of 18 to a minimum of 0, with higher scores representing higher risk of death. A score of five, the highest in one category, was assigned to the inability to follow commands. A score of four was assigned to systolic blood pressure between 0 and 49, which was the second-highest score in one category (**Table 2**).

Validation of the Models

The AUROC analysis showed that the neural network models had excellent performance in both the internal testing set and the validation set (internal testing set: **Supplementary Table 2**, **Supplementary Figure 1**; validation set: **Table 3**, **Figure 2**). NN-GAPSO showed the highest performance using the total GCS. NN-CAPSO replaced the initial GCS with the assessment of whether the GCS-m was <6, and its AUROC was lower than that of NN-GAPSO ($p < 0.001$). After further removal of systolic blood pressure, the AUROC values of NN-CAPO decreased in comparison to NN-CAPSO ($p < 0.001$). The AUROC of the CAPSO score was lower than that of NN-GAPSO and NN-CAPSO (both $p < 0.001$) but the same as that of NN-CAPO ($p > 0.05$). The AUROCs of NN-GAPSO and NN-CAPSO were higher than those of other scores (except TRISS), such as RTS, NTS, GAP, MGAP, and TRIGAGES (all $p < 0.001$). The AUROCs of NN-CAPO and CAPSO scores were similar to that of TRIAGES (both $p > 0.05$) and higher than the rest of the above scores (all $p < 0.001$). The sensitivity, specificity and accuracy of the models and scores according to the sensitivity closest to 0.95 are shown in **Table 3**. The probability calibration curves in the validation set of the neural network models, CAPSO score, and other trauma scores calibrated with the training set are shown in **Figure 3**, while those in the internal testing set are shown in **Supplementary Figure 2**. The SHAP plots of NN-CAPSO model are shown in **Supplementary Figure 3**.

Table 4 shows the reclassification of NN-CAPSO with NN-GAPSO, NN-CAPO, CAPSO score, TRIAGES score, and TRISS score in the validation set for the severity of trauma in 100,000 randomly selected patients. In NN-CAPSO, compared with NN-GAPSO, a total of 6,163 patients were misclassified out of 100,000 patients, of which 3,648 were overtriaged and the remaining 2,515 were undertriaged. Compared with NN-CAPSO, NN-CAPO misclassified 4,849 patients (overtriaged: 2,003, undertriaged: 2,846), while the CAPSO score misclassified 5,170 patients (overtriaged: 1,217, undertriaged: 3,953). When compared with TRIAGES, NN-CAPSO correctly reclassified 8,612 patients into the intermediate-risk group, which was classified by TRIAGES into the low-risk group, and correctly reclassified 344 patients into the high-risk group, which was classified by TRIAGES into the intermediate-risk group. However, in this comparison, NN-CAPSO incorrectly classified 326 patients who should have been in the high-risk group into the intermediate-risk group, and 927 patients who should have been in the intermediate-risk group were incorrectly classified into the low-risk group. Compared with TRISS, NN-CAPSO correctly classified 11,649 patients in the intermediate-risk group, who were classified as low risk by TRISS, but incorrectly classified 1,251 patients in the high-risk group as belonging to the intermediate-risk group.

TABLE 3 | Comparison of the diagnostic properties of the models/scores at a sensitivity threshold of nearest 95%.

Models/Scores	Variables	AUROC [95% CI]	Sensitivity [95% CI]	Specificity [95% CI]	Accuracy [95% CI]
NN-GAPSO	GCS, Age, Pulse rate, SBP, SpO_2	0.921 [0.918, 0.923]	0.951 [0.949, 0.954]	0.559 [0.557, 0.560]	0.578 [0.577, 0.580]
NN-CAPSO	Ability to follow commands, Age, Pulse rate, SBP, SpO_2	0.911 [0.909, 0.913]	0.951 [0.948, 0.954]	0.546 [0.545, 0.548]	0.566 [0.565, 0.568]
NN-CAPO	Ability to follow commands, Age, Pulse rate, SpO_2	0.904 [0.902, 0.906]	0.951 [0.948, 0.954]	0.518 [0.517, 0.520]	0.540 [0.538, 0.541]
CAPSO	Ability to follow commands, Age, Pulse rate, SBP, SpO_2	0.904 [0.902, 0.906]	0.960 [0.957, 0.963]	0.492 [0.490, 0.493]	0.515 [0.513, 0.516]
RTS	SBP, RR, GCS	0.851 [0.848, 0.854]	0.760 [0.754, 0.765]	0.879 [0.878, 0.880]	0.873 [0.872, 0.874]
NTS	SBP, SpO_2, GCS	0.888 [0.885, 0.891]	0.938 [0.935, 0.942]	0.391 [0.390, 0.393]	0.418 [0.417, 0.420]
MGAP	Mechanism, GCS, Age, SBP	0.898 [0.896, 0.901]	0.952 [0.949, 0.955]	0.451 [0.449, 0.452]	0.476 [0.474, 0.477]
GAP	GCS, Age, SBP	0.897 [0.894, 0.899]	0.967 [0.965, 0.970]	0.377 [0.375, 0.378]	0.406 [0.404, 0.407]
TRIAGES	GCS, Age, SBP, RR	0.903 [0.900, 0.905]	0.976 [0.974, 0.978]	0.349 [0.347, 0.350]	0.380 [0.378, 0.381]
TRISS	Mechanism, GCS, Age, SBP, RR, ISS	0.934 [0.932, 0.936]	0.959 [0.956, 0.961]	0.575 [0.573, 0.576]	0.594 [0.592, 0.595]

CI, confidence interval; NN, Neural network; GAPSO, Glasgow Coma Scale, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; CAPSO, the Ability to Follow Commands, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; CAPO, the Ability to Follow Commands, Age, Pulse Rate, and Peripheral Oxygen saturation; RTS, Revised Trauma Score; NTS, New Trauma Score; MGAP, Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure; GAP, Glasgow Coma Scale, Age, and Systolic Blood Pressure score; TRIAGES, Trauma Rating Index in Age, Glasgow Coma Scale, Respiratory rate and Systolic blood pressure; TRISS, Trauma and Injury Severity Score; AUROC, Area Under the Receiver Operating Characteristics; GCS, Glasgow Coma Scale; SBP, Systolic Blood Pressure; RR, Respiratory rate; SpO_2, Peripheral Oxygen saturation; ISS, Injury Severity Score.

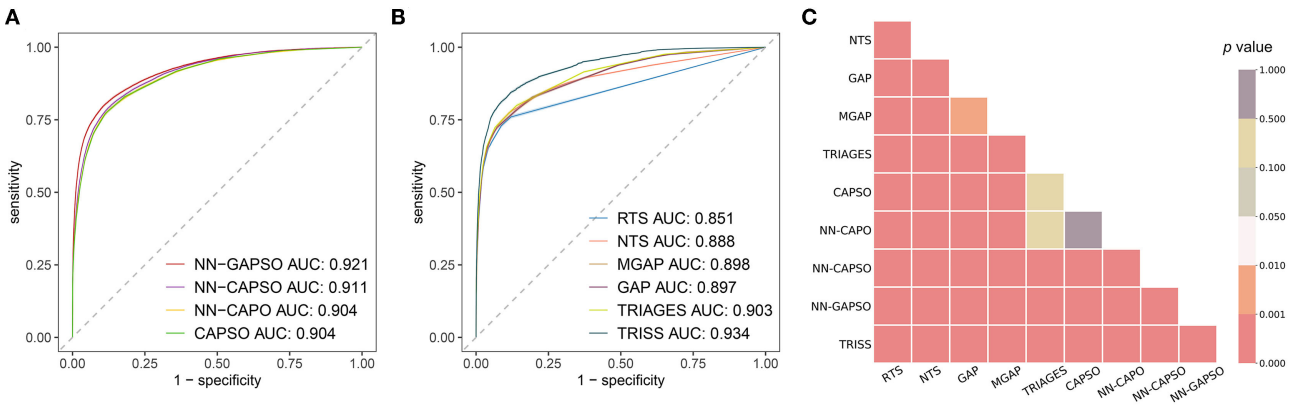


FIGURE 2 | The discrimination of models/scores in the validation cohort. **(A)** Receiver operating characteristic curves for newly developed models; **(B)** receiver operating characteristic curves for trauma scores; **(C)** *p* values for a two-by-two comparison between different models and scores. *NN*, Neural network; *GAPSO*, Glasgow Coma Scale, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; *CAPSO*, the Ability to Follow Commands, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; *CAPO*, the Ability to Follow Commands, Age, Pulse Rate, and Peripheral Oxygen saturation; *RTS*, Revised Trauma Score; *NTS*, New Trauma Score; *MGAP*, Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure; *GAP*, Glasgow Coma Scale, Age, and Systolic Blood Pressure score; *TRIAGES*, Trauma Rating Index in Age, Glasgow Coma Scale, Respiratory rate and Systolic blood pressure; *TRISS*, Trauma and Injury Severity Score.

DISCUSSION

The aim of this study was to develop a trauma mortality prediction model using a simple binary assessment of GCS-Motor (GCS-m) score <6, namely, whether the patient has the ability to follow commands, instead of the GCS. The prediction accuracy of the neural network-based CAPSO model was still higher than that of the other prehospital trauma scores using the total GCS, although it was slightly worse than that of the neural network-based GAPSO model, which uses the total GCS. In addition, the logistic regression-based CAPSO score had a predictive power similar to that of the TRIAGES score, and it was superior to other prehospital trauma scores. In addition, the neural network model NN-CAPO, which used the assessment of GCS-m <6 and removed the variable SBP, could achieve predictive accuracy similar to that of the TRIAGES.

In this study, the cutoff value for predicting the probability of death was chosen first based on the sensitivity closest to 95%, referring to a previous study (5). However, higher sensitivity tends to be accompanied by lower specificity, and in this dataset, even TRISS failed to reach the upper 60% of specificity. In addition, the reclassification table based on the classification of trauma patients according to minor, moderate, and severe injuries is more relevant for practical use. Furthermore, in this study, to make the results more intuitive, a random sample of 100,000 patients in the validation set was selected and

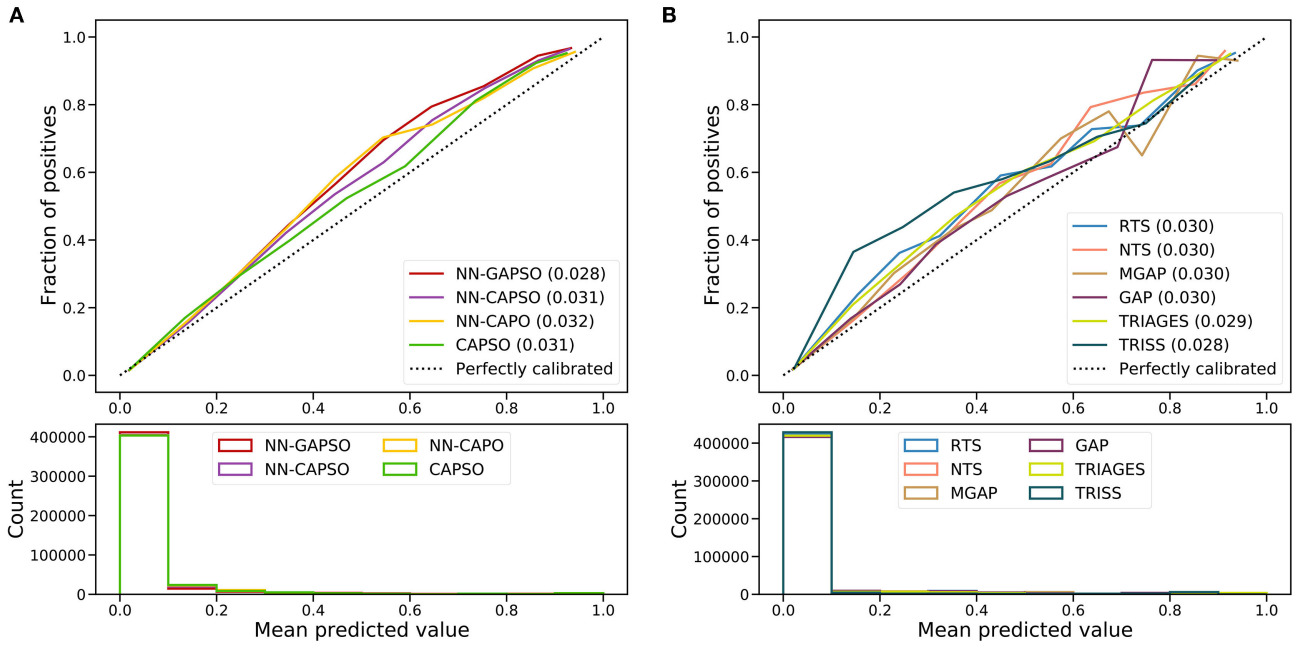


FIGURE 3 | Calibration curves of newly developed models **(A)** and trauma scores **(B)** in the validation cohort. *NN*, Neural network; *GAPSO*, Glasgow Coma Scale, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; *CAPSO*, the Ability to Follow Commands, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; *CAPO*, the Ability to Follow Commands, Age, Pulse Rate, and Peripheral Oxygen saturation; *RTS*, Revised Trauma Score; *NTS*, New Trauma Score; *MGAP*, Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure; *GAP*, Glasgow Coma Scale, Age, and Systolic Blood Pressure Score; *TRIAGES*, Trauma Rating Index in Age, Glasgow Coma Scale, Respiratory rate and Systolic blood pressure; *TRISS*, Trauma and Injury Severity Score.

grouped according to the 5% and 50% cutoff values of predicted mortality, referring to previous studies (5, 6). According to the reclassification table, NN-CAPSO could correctly triage more patients with moderate and severe injuries than the TRIAGES score.

In recent years, with the rise of machine learning algorithms, there has been an increasing number of studies using machine learning methods other than logistic regression algorithms to build prediction models. Most of these studies have suggested that machine learning algorithms have satisfying performance and broad application prospects in the medical field (23). Nevertheless, skepticism is also present. It was concluded that no evidence was found that the machine learning algorithms outperformed the logistic regression algorithm (24). In low-dimensional data, the machine learning algorithms were not considered to perform better than logistic regression (25). Some researchers have pointed out that the advantage of machine learning algorithms comes into play when dealing with data with a large number of features (26, 27), while others have claimed that machine learning algorithms require a larger data volume to demonstrate their performance (28). In addition to data quantity and dimensionality, the nature and processing of the features also play a very important role when comparing algorithms, such as the processing of continuous variables and the generation of interaction terms. In this study, although the number of features was relatively small, there was a sufficient amount of data, and the neural network models outperformed the logistic regression model and scores. Currently, with the widespread availability of smart electronic devices, machine learning models for predicting the outcomes of trauma patients, embedded into applications, will have higher accuracy and efficiency compared to scores calculated manually, whether applied for rapid assessment of trauma patient severity in normal times or in MCI. However, it is difficult to make machine learning algorithms interpretable, especially neural network algorithms, which are often referred to as "black boxes." In this study, SHAP values are used to interpret the neural network model NN-CAPSO. Furthermore, machine learning algorithms are still not as commonly used in practice as intuitive scoring systems. Therefore, in addition to neural network algorithms, we have developed a simple scoring system for CAPSO using the logistic regression algorithm to facilitate the validation and use of this system in clinical settings.

Currently, pulse rate and peripheral oxygen saturation are easy to obtain in the prehospital phase. The accuracy of peripheral oxygen saturation measurement is relatively reliable when it is above 75% (29). In the CAPSO scoring system, we set the threshold for peripheral oxygen saturation at 80% according to the coefficients of the multivariate generalized additive model. In the SHAP plot, age was the second most important feature in the ranking. Age was also included as a variable in the MGAP, GAP, and TRIAGES scores which performed well in ROC analysis. By entering the age into the intelligent device in advance,

TABLE 4 | Reclassification of severity between NN-CAPSO and other scoring systems in the randomly selected validation cohort[a].

Reclassification of severity between NN-CAPSO and NN-GAPSO					
		NN-GAPSO			
Scores	**Severity**	**Mild (<0.05 points)**	**Moderate (0.05 to 0.5 points)**	**Severe (>0.5 points)**	**Total**
NN-CAPSO	Mild (<0.05 points)	80,483 (0.96)	2,034 (6.64)	0 (0.0)	82,517 (1.1)
	Moderate (0.05 to 0.5 points)	3,500 (3.51)	11,921 (18.25)	481 (71.1)	15,902 (16.61)
	Severe (>0.5 points)	1 (0.0)	148 (36.49)	1,432 (90.92)	1,581 (85.77)
	Total	83,984 (1.06)	14,103 (16.77)	1,913 (85.94)	100,000 (4.9)
Reclassification of severity between NN-CAPSO and NN-CAPO					
		NN-CAPO			
Scores	**Severity**	**Mild (<0.05 points)**	**Moderate (0.05 to 0.5 points)**	**Severe (>0.5 points)**	**Total**
NN-CAPSO	Mild (<0.05 points)	80,675 (1.03)	1,842 (4.13)	0 (0.0)	82,517 (1.1)
	Moderate (0.05 to 0.5 points)	2,482 (6.08)	13,259 (18.06)	161 (59.63)	15,902 (16.61)
	Severe (>0.5 points)	0 (0.0)	364 (68.96)	1,217 (90.8)	1,581 (85.77)
	Total	83,157 (1.18)	15,465 (17.59)	1,378 (87.16)	100,000 (4.9)
Reclassification of severity between NN-CAPSO and CAPSO					
		CAPSO			
Scores	**Severity**	**Mild (<5 points)**	**Moderate (5 to 10 points)**	**Severe (>10 points)**	**Total**
NN-CAPSO	Mild (<0.05 points)	81,461 (1.06)	1,056 (4.17)	0 (0.0)	82,517 (1.1)
	Moderate (0.05 to 0.5 points)	3,708 (6.07)	12,033 (19.35)	161 (54.66)	15,902 (16.61)
	Severe (>0.5 points)	0 (0.0)	245 (67.35)	1336 (89.15)	1581 (85.77)
	Total	85,169 (1.28)	13,334 (19.03)	1,497 (85.44)	100,000 (4.9)
Reclassification of severity between NN-CAPSO and TRIAGES					
		TRIAGES			
Scores	**Severity**	**Mild (<4 points)**	**Moderate (5 to 8 points)**	**Severe (>9 points)**	**Total**
NN-CAPSO	Mild (<0.05 points)	81,590 (1.0)	927 (9.39)	0 (0.0)	82,517 (1.1)
	Moderate (0.05 to 0.5 points)	8,612 (6.08)	6,964 (27.31)	326 (65.95)	15,902 (16.61)
	Severe (>0.5 points)	18 (38.89)	344 (61.05)	1,219 (93.44)	1,581 (85.77)
	Total	90,220 (1.5)	8,235 (26.7)	1,545 (87.64)	100,000 (4.9)
Reclassification of severity between NN-CAPSO and TRISS					
		TRISS			
Scores	**Severity**	**Mild(>0.834 points)**	**Moderate (0.353 to 0.834 points)**	**Severe(<0.353 points)**	**Total**
NN-CAPSO	Mild (<0.05 points)	82,029 (0.98)	465 (20.22)	23 (34.78)	82,517 (1.1)
	Moderate (0.05 to 0.5 points)	11,649 (6.3)	3,002 (35.04)	1,251 (68.35)	15,902 (16.61)
	Severe (>0.5 points)	83 (36.14)	232 (68.53)	1,266 (92.18)	1,581 (85.77)
	Total	93,761 (1.67)	3,699 (35.28)	2,540 (79.92)	100,000 (4.9)

[a]*Data in parentheses are the percentages of deaths (%). Severe, high risk (> 50%) of death; Moderate, intermediate risk of death; Mild, low risk (< 5%) of death. NN, Neural network; GAPSO, Glasgow Coma Scale, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; CAPSO, the Ability to Follow Commands, Age, Pulse Rate, Systolic Blood Pressure, and Peripheral Oxygen saturation; CAPO, the Ability to Follow Commands, Age, Pulse Rate, and Peripheral Oxygen saturation; RTS, Revised Trauma Score; NTS, New Trauma Score; MGAP, Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure; GAP, Glasgow Coma Scale, Age, and Systolic Blood Pressure score; TRIAGES, Trauma Rating Index in Age, Glasgow Coma Scale, Respiratory rate and Systolic blood pressure; TRISS, Trauma and Injury Severity Score.*

it ensures that the age is available first when assessing the severity of the patient's injury by models or scores, which is applicable to people who wear the device earlier, such as military personnel or firefighters. However, if the patient is unconscious and the age is not available from all other sources in a short period of time, guesses by medical personnel can be useful but may lead to some degree of degradation in model accuracy, which is still subject to further validation. For the assessment of the patient's

state of consciousness, the GCS is mostly used today. However, in specific situations, such as MCI, complete measurement of the GCS will waste precious time, as it has been reported that even formally trained clinicians have a probability of up to 20% of making errors in assessing GCS in a normal setting (30), let alone in the complicated trauma field. It was reported that the motor component of the GCS not only correlated linearly with survival but also retained most of the predictive validity of the GCS (31). When using GCS-m<6 as a predictor for the need for treatment at a trauma center, this predictor showed comparable validity to that using total GCS $\leq$ 13 (32). In addition, it was difficult to take manual measurements of blood pressure in the field (14). In this study, we attempted to replace the GCS with the ability to follow commands and further to remove SBP to build models and compare the effect of different models on the classification results. NN-GAPSO reached the highest AUROC as expected, while the performance of NN-CAPSO, NN-CAPO, and CAPSO deteriorated when compared to NN-GAPSO, but not as much as predicted. When the severity of the randomly selected patients was reclassified according to three intervals of mortality, the NN-GAPSO and NN-CAPSO disagreed on a total of 6,164 patients (6.164%, AUROC difference was nearly 0.01), while NN-CAPSO and CAPSO score had a different classification for a total of 5,170 patients (5.17%, AUROC difference was nearly 0.007). The slight sacrifice of models' performance in exchange for more ease of application was considered to make sense, even the accuracy of simplified models was not weaker than that of the other scores applicable to prehospital settings. The employment of the simpler model is estimated to increase user-friendliness and improve the efficiency of triage, although it remains to be evaluated in other datasets or in a real field setting. Furthermore, vital signs, assessment of consciousness, and age data of trauma patients may be missing due to specific comorbidities, injury conditions, or treatments. Despite the population with missing values is not very large, they may benefit from specific models developed for them. Alternatively, the use of algorithms that are able to handle missing values, or build models that treat missing values as special values, can preserve the information of the missing values themselves and facilitate the application to trauma patients with incomplete information, which requires further research.

In recent years, various emerging technologies are bringing about changes in the method of triage. The Wireless Vital Signs Monitor (WVSM) is a wireless vital sign monitoring device, and with its help, a health care worker can monitor up to 20 patients at the same time, making it very suitable for triaging in the field (33). Moreover, the use of smart glasses for remote classification is promising in reaching high accuracy, either through algorithms embedded into the glasses or by remote video connection to other physicians (34, 35). The use of wearable devices or radar for remote vital sign monitoring was supposed to save considerable manpower and time (36). However, it is complicated to apply GCS scores for consciousness assessment in the remote situation. In this case, the application of binary assessment of GCS-m score <6 rather than GCS would be effective. For example, some instructions from corresponding devices will ask the casualty to complete certain actions, and feedback can then be input into the devices to determine whether the person has the ability to follow commands. Then, the scoring model embedded in devices would give advice on triage. Furthermore, it is not easy to measure SBP with lightweight wearable devices or radar. Therefore, using machine learning algorithms to develop triage models not involving SBP will also be highly applicable now and in the near future.

In summary, the new user-friendly CAPSO system makes it possible to rapidly and reliably predict in-hospital mortality in trauma patients. It is suitable for future prehospital intelligent automated triage applications and is expected to improve the efficiency of triage by integration into prehospital decision-making systems.

LIMITATIONS

This study used only data from the NTDB database and therefore has the limitations of that database. A portion of patients with much missing information were excluded, but the absence of vital signs may be due to the patient being agitated or receiving emergency medical care, etc., so the final study population included may have been to a degree biased, although the large sample population of this study is likely to be useful in reducing bias. While the data from 2017 were used separately for validation, the models or scores created for this study still require validation with data from other sources, particularly prospective data. For comparison with other studies, the cut-off values for the probability selected for this study referenced previous studies; however, in practice, the results of the model or score should be corrected and the appropriate cut-off values should be selected for decision making based on the application scenario.

DATA AVAILABILITY STATEMENT

The datasets analyzed for this study are available for purchase from the American College of Surgeons (ACS) via https://www.facs.org/quality-programs/trauma/tqp/center-programs/ntdb/datasets.

ETHICS STATEMENT

All data provided by the NTDB are de-identified. Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YLi designed the study, analyzed the data, and drafted the manuscript. LW and YLiu contributed to the acquisition of data and conducted data cleaning. YF, RY, and MY analyzed and interpreted the data. YZ, FZ, and HK jointly conceived of and designed this study. ZZ and HK conducted critical revision

of the article. All of the authors reviewed and approved the final manuscript.

FUNDING

This work was supported by the National Key Research and Development Program of China (2020YFB1313901) and Research and Development Project of Medical Big Data and Artificial Intelligence in PLA General Hospital (2019MBD-014).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.810195/full#supplementary-material

REFERENCES

1. Polinder S, Haagsma JA, Toet H, Beeck E. Epidemiological burden of minor, major and fatal trauma in a national injury pyramid. *Br J Surg.* (2011) 99(S1):114–21. doi: 10.1002/bjs.7708
2. Wright CA. A National Evaluation of the Effect of Trauma-center Care on mortality. *J Trauma Nurs.* (2006) 13:366. doi: 10.1097/00043860-200607000-00018
3. Kluger Y, Coccolini F, Catena F, Ansaloni L. *WSES Handbook of Mass Casualties Incidents Management*: Cham: Springer (2020).
4. Champion HR, Sacco WJ, Copes WS, Gann DS, Gennarelli TA, Flanagan ME, et al. A revision of the Trauma Score. *J Trauma.* (1989) 29:623–9. doi: 10.1097/00005373-198905000-00017
5. Sartorius D, Le Manach Y, David JS, Rancurel E, Smail N, Thicoïpé M, et al. Mechanism, Glasgow Coma Scale, Age, and Arterial Pressure (MGAP): a new simple prehospital triage score to predict mortality in trauma patients. *Crit Care Med.* (2010) 38:831–7. doi: 10.1097/CCM.0b013e3181cc4a67
6. Kondo Y, Abe T, Kohshi K, Tokuda Y, Cook EF, Kukita I. Revised trauma scoring system to predict in-hospital mortality in the emergency department: Glasgow coma scale, age, and systolic blood pressure score. *Crit Care.* (2011) 15:R191. doi: 10.1186/cc10348
7. Jeong JH, Park YJ, Kim DH, Kim TY, Kang C, Lee SH, et al. The new trauma score (NTS): a modification of the revised trauma score for better trauma mortality prediction. *BMC Surg.* (2017) 17:77. doi: 10.1186/s12893-017-0272-4
8. Shiraishi A, Otomo Y, Yoshikawa S, Morishita K, Roberts I, Matsui H. Derivation and validation of an easy-to-compute trauma score that improves prognostication of mortality or the Trauma Rating Index in Age, Glasgow Coma Scale, Respiratory rate and Systolic blood pressure (TRIAGES) score. *Crit Care.* (2019) 23:365. doi: 10.1186/s13054-019-2636-x
9. Baker SP. o'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma Acute Care Surg.* (1974) 14:187–96. doi: 10.1097/00005373-197403000-00001
10. Boyd CR, Tolson MA, Copes WS. Evaluating trauma care: the TRISS method. Trauma Score and the Injury Severity Score. *J Trauma.* (1987) 27:370–8. doi: 10.1097/00005373-198704000-00005
11. Reith FC, Van den Brande R, Synnot A, Gruen R, Maas AI. The reliability of the Glasgow Coma Scale: a systematic review. *Intensive Care Med.* (2016) 42:3–15. doi: 10.1007/s00134-015-4124-3
12. Riechers RG, Ramage A, Brown W, Kalehua A, Rhee P, Ecklund JM, et al. Physician knowledge of the Glasgow Coma Scale. *J Neurotrauma.* (2005) 22:1327–34. doi: 10.1089/neu.2005.22.1327
13. Chou R, Totten AM, Carney N, Dandy S, Fu R, Grusing S, et al. Predictive utility of the total Glasgow Coma Scale Versus the Motor Component of the Glasgow Coma Scale for Identification of patients with serious Traumatic injuries. *Ann Emerg Med.* (2017) 70:143–57.e6. doi: 10.1016/j.annemergmed.2016.11.032
14. McManus J, Yershov AL, Ludwig D, Holcomb JB, Salinas J, Dubick MA, et al. Radial pulse character relationships to systolic blood pressure andtrauma outcomes. *Prehosp Emerg Care.* (2005) 9:423–8. doi: 10.1080/10903120500255891
15. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. *Crit Care Med.* (2016) 44:368. doi: 10.1097/CCM.0000000000001571
16. Kupas DF, Melnychuk EM, Young AJ. Glasgow Coma Scale motor component ("Patient Does Not Follow Commands") performs similarly to total Glasgow Coma Scale in predicting severe injury in Trauma patients. *Ann Emerg Med.* (2016) 68:744–50.e3. doi: 10.1016/j.annemergmed.2016.06.017
17. Hashmi ZG, Kaji AH, Nathens AB. Practical guide to surgical data sets: National Trauma Data Bank (NTDB). *JAMA Surg.* (2018) 153:852–3. doi: 10.1001/jamasurg.2018.0483
18. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg.* (2015) 102:148–58. doi: 10.1002/bjs.9736
19. Buuren SV, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. *J Stat Softw.* (2011) 45:1–67. doi: 10.18637/jss.v045.i03
20. Kriegeskorte N, Golan T. Neural network models and deep learning. *Current Biology.* (2019) 29: R231–6. doi: 10.1016/j.cub.2019.02.034
21. Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* (1988) 44:837. doi: 10.2307/2531595
22. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell.* (2020) 2:56–67. doi: 10.1038/s42256-019-0138-9
23. Beam AL, Kohane IS. Big data and machine learning in health care. *JAMA.* (2018) 319:1317–8. doi: 10.1001/jama.2017.18391
24. Ec A, Jie MB, Gscb C, Ews D, Jyvae F, Bvca D, et al. systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* (2019) 110:12–22. doi: 10.1016/j.jclinepi.2019.02.004
25. Gravesteijn BY, Nieboer D, Ercole A, Lingsma HF, Zoerle T. Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury. *J Clin Epidemiol.* (2020) 122:95–107. doi: 10.1016/j.jclinepi.2020.03.005
26. Deo RC, Nallamothu BK. Learning about machine learning: the promise and pitfalls of big data and the electronic health record. *Am Heart Assoc.* (2016) 9:618–20. doi: 10.1161/CIRCOUTCOMES.116.003308
27. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning for electronic health records. *NPJ Digit Med.* (2018) 1:18. doi: 10.1038/s41746-018-0029-1
28. Tjeerd V, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Med Res Methodol.* (2014) 14:137. doi: 10.1186/1471-2288-14-137
29. Wouters PF, Gehring H, Avgerinos J, Konecny E, Meyfroidt G. Accuracy of pulse oximeters: the European multi-center trial. *Anesth Analg.* (2002) 94:S13–S6.
30. Holt AW, Bury LK, Bersten AD, Skowronski GA, Vedig AE. Prospective evaluation of residents and nurses as severity score data collectors. *Crit Care Med.* (1992) 20:1688–91. doi: 10.1097/00003246-199212000-00015
31. Healey C, Osler TM, Rogers FB, Healey MA, Glan Ce LG, Kilgo PD, et al. Improving the Glasgow Coma Scale score: motor score alone is a better predictor. *J Trauma.* (2003) 54:671–8. doi: 10.1097/01.TA.0000058130.30490.5D
32. Brown JB, Forsythe RM, Stassen NA, Peitzman AB, Gestring ML. Evidence-based improvement of the National Trauma Triage Protocol: the Glasgow

Coma Scale versus Glasgow Coma Scale motor subscale. *J Trauma Acute Care Surg.* (2014) 77:101–2. doi: 10.1097/TA.0000000000000280
33. Salinas J, Nguyen R, Darrah MI, Kramer GA, Cancio LC. Advanced monitoring and decision support for battlefield critical care environment. *US Army Med Dep J.* (2011):73–81.
34. Follmann A, Ohligs M, Hochhausen N, Beckers SK, Rossaint R, Czaplik M. Technical support by smart glasses during a mass casualty incident: a randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine. *J Med Internet Res.* (2019) 21:e11939. doi: 10.2196/11939
35. Mccoy E, Alrabah R, Weichmann W, Langdorf MI, Lotfipour S. Feasibility of telesimulation and google glass for mass casualty triage education and training. *West J Emerg Med.* (2019) 20:512–9. doi: 10.5811/westjem.2019.3.40805
36. Kim D, You S, So S, Lee J, Yook S, Jang DP, et al. A data-driven artificial intelligence model for remote triage in the prehospital environment. *PLoS ONE.* (2018) 13:e0206006. doi: 10.1371/journal.pone.0206006

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fmed.2021.792974

Application of Machine Learning to Predict Acute Kidney Disease in Patients With Sepsis Associated Acute Kidney Injury

Jiawei He†, *Jin Lin*† *and Meili Duan**

Department of Critical Care Medicine, Beijing Friendship Hospital, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Saman Sarraf,
Institute of Electrical and Electronics Engineers, United States
Kasem Khalil,
Western Kentucky University, United States

***Correspondence:**
Meili Duan
dmeili@ccmu.edu.cn

†*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *11 October 2021*
Accepted: *08 November 2021*
Published: *10 December 2021*

Citation:
He J, Lin J and Duan M (2021) Application of Machine Learning to Predict Acute Kidney Disease in Patients With Sepsis Associated Acute Kidney Injury. Front. Med. 8:792974. doi: 10.3389/fmed.2021.792974

Background: Sepsis-associated acute kidney injury (AKI) is frequent in patients admitted to intensive care units (ICU) and may contribute to adverse short-term and long-term outcomes. Acute kidney disease (AKD) reflects the adverse events developing after AKI. We aimed to develop and validate machine learning models to predict the occurrence of AKD in patients with sepsis-associated AKI.

Methods: Using clinical data from patients with sepsis in the ICU at Beijing Friendship Hospital (BFH), we studied whether the following three machine learning models could predict the occurrence of AKD using demographic, laboratory, and other related variables: Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM), decision trees, and logistic regression. In addition, we externally validated the results in the Medical Information Mart for Intensive Care III (MIMIC III) database. The outcome was the diagnosis of AKD when defined as AKI prolonged for 7–90 days according to Acute Disease Quality Initiative-16.

Results: In this study, 209 patients from BFH were included, with 55.5% of them diagnosed as having AKD. Furthermore, 509 patients were included from the MIMIC III database, of which 46.4% were diagnosed as having AKD. Applying machine learning could successfully achieve very high accuracy (RNN-LSTM AUROC = 1; decision trees AUROC = 0.954; logistic regression AUROC = 0.728), with RNN-LSTM showing the best results. Further analyses revealed that the change of non-renal Sequential Organ Failure Assessment (SOFA) score between the 1st day and 3rd day (Δnon-renal SOFA) is instrumental in predicting the occurrence of AKD.

Conclusion: Our results showed that machine learning, particularly RNN-LSTM, can accurately predict AKD occurrence. In addition, Δ $SOFA_{non-renal}$ plays an important role in predicting the occurrence of AKD.

Keywords: intensive care unit, sepsis, acute kidney injury, acute kidney disease, machine learning

INTRODUCTION

The prevalence of acute kidney injury (AKI) in patients admitted to intensive care units (ICU) is approximately 50%. Nearly half of all AKI cases are present with sepsis, which may further worsen the prognosis (1, 2). Previous studies have reported the mortality rate of ICU patients with septic AKI as 30–45%, with the survivors still associated with the increased risk of chronic kidney disease (CKD) and cardiovascular events (3).

Increased severity and higher duration of AKI are associated with poor prognosis. In line with several previous results, Kellum et al. reported poorer clinical outcomes in patients with AKI lasting longer than 7 days than in patients who had renal function recovered within 7 days (4). Similar results have been previously reported in other studies (5, 6). Furthermore, in patients who developed sepsis persistent AKI beyond 7 days was associated with adverse clinical outcomes (5, 6). Hence, Acute Disease Quality Initiative-16 (ADQI-16) workshop suggested defining acute kidney disease (AKD) as impaired kidney function lasting 7–90 days after AKI (7). Unlike AKI patients, whose renal function typically recovers within 7 days, AKD patients suffer from persistent renal impairment and often have poor clinical outcomes (8).

Recent studies have utilized machine learning techniques for predicting AKI. Using machine learning techniques such as logistic regression and extreme gradient boosting (XGBoost), Zhang et al. identified some important clinical factors associated with AKI such as age, urinary creatinine concentration, maximum blood urea nitrogen concentration, and albumin (9). Zimmerman et al. showed that comprehensive demographics and physiologic features can accurately predict max serum creatinine level during day 2 and day 3 and also predict new AKI onset by cross-validation on linear regression and multiple machine learning models (10). However, AKD prediction has not been reported.

The AKD phase is a time window for potentially initiating key interventions to alter the natural history of kidney disease (7), and thus, the early identification of patients at high risk of developing AKD is important. Previous studies have shown that age, hypertension, diabetes mellitus, the history of CKD, the severity of AKI, and the use of mechanical ventilators were associated with the onset of AKD (11–17), however, machine learning methods have been seldom used to predict the occurrence of AKD. This study was aimed at using longitudinal data to predict the occurrence of AKD.

MATERIALS AND METHODS

Data Source and Participants

Patients were recruited from the intensive care unit of Beijing Friendship Hospital (BFH), between January 1, 2015 and December 21, 2020. We obtained electronic healthcare data from Medical Information Mart for Intensive Care III (MIMIC III) (18). The inclusion criteria were as follows: (1) age $\geq$ 18 years old; (2) AKI caused by sepsis. The exclusion criteria were as follows: (1) AKI duration <48 h; (2) length of survival time <7 days; (3) CKD stage 5 or end-stage kidney disease defined as estimated glomerular filtration rate <15 ml/min/1.73 m^2; (4) patients with missing important data (e.g., data on demographics and variables for calculating traditional severity scores). The study was reported according to the recommendations of the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement (19).

Data Extraction

We extracted the following data from BFH and the MIMIC III records upon admission to ICU (day 1): (1) demographic information; (2) ICU details, including vitals, laboratory data, mechanical ventilation requirement, and exposure to nephrotoxic drugs; (3) severity of illness was measured using Simplified Acute Physiology Score II (SAPS II), Acute Physiological Score III (APS III), and non-renal Sequential Organ Failure Assessment (SOFA) score. The data on non-renal SOFA, creatinine, and urine output were recorded daily until day 3. Delta non-renal SOFA, delta creatinine, and delta urine output was the difference between the value at day 3 and the admission value.

Outcomes and Definitions

The occurrence of AKD was the primary outcome. AKD was defined as the presentation of at least KDIGO Stage 1 criteria for >7 days after an AKI-initiating event, which agrees with the diagnostic criteria proposed by ADQI-16 in 2017 (7). The definition of sepsis was based on the diagnostic criteria of the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3), including a suspected infection and a SOFA score of $\geq$2 (20). The Kidney Disease: Improving Global Outcomes (KDIGO) classification according to both serum creatinine (SCr) and urine output (UO) criteria were used to define AKI (21). CKD was defined according to the Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease (22).

Sample Size

The sample size was defined as having at least 10 outcome events per variable per estimated parameter according to a previous study (23). Our sample and the number of AKD approached that determined by the calculated result.

Statistical Analysis

Values were presented as total numbers (percentages) for categorical variables and the means $\pm$ SDs or medians (interquartile ranges) for continuous variables. Comparisons were made using the Student's *t*-test or rank-sum test for continuous variables, and the Chi-square test or Fisher's exact test for categorical variables, as appropriate. All statistical tests were two-sided, and *P*-values of <0.05 were considered statistically significant.

Model Development and Validation

The included patients from BFH and MIMIC III comprised the training dataset and the validation dataset, respectively. We selected three models for comparison: Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM), decision tree, and logistic regression. The discrimination performance of these models in the training dataset and the validation

dataset was evaluated by area under the receiver operating characteristic (AUROC).

Recurrent Neural Network-Long Short-Term Memory

The RNN has been widely used to handle the longitudinal variables, LSTM is one type of RNN (24, 25). It can effectively process a large amount of sequential data. It comprises several modules, which can store the processed data from the previous stage. Unlike ordinary RNN, classic LSTM comprises several modules called cells. Data can be transferred from the previous cell to the next cell, including input gate, forget gate, and output gate. All data are added to the input gate, and the output gate displays the final data result. Unlike ordinary RNN which can have only one memory stacking method, LSTM can control the transmission state through the gating state, remember important information and forget unimportant information. The forget gate can enhance the ability of LSTM to process data and avoid the problem of data dependence.

Decision Tree

Decision tree/random forest can predict the classification (AKD or non-AKD) from the data, which can display the decision result more clearly (26). We can use the decision tree to interpret the prediction results. The process from the root to the leaf of the tree shows the prediction classification, according to the algorithm of the decision tree. Each step of the decision tree involves checking a piece of data. If the predictor satisfied a certain condition, it would follow the upper branch to indicate type 0, predicting that AKD will occur. Otherwise, it would follow the lower branch to indicate type 1, predicting that AKD will not occur. The decision trees were trained to create a model that could factor in multiple input variables and predict the value of the target variable. The division of the tree continues until the node contains the minimum number of training examples or reaches the maximum tree depth. The complexity parameter is used to indicate the prediction performance, which depends on how many classes are mixed in the two groups generated by the decision tree (27). We choose the number of leaves when the complexity parameter is the lowest to minimize the chance of making errors in the decision tree.

Logistic Regression

In the training dataset, we used the Least Absolute Shrinkage and Selection Operator (LASSO) method to select the most useful predictive variables (28). Continuous variables were made into dichotomous variables and were entered into a logistic regression with other variables. The nomogram predicting the occurrence of AKD was established using the LASSO method for the selected variables. The performance of the nomogram was evaluated by calibration curves. The calibration evaluation uses a calibration chart to show the relationship between the observed frequency and the predicted probability. The nomogram was verified in the validation dataset to evaluate the stability of the nomogram. In addition, decision curve analysis (DCA) was used to evaluate the clinical utility of the final nomogram (29). The net benefit is calculated by subtracting the proportion of false positives from true positives (30).

Moreover, the discrimination of three machine learning algorithms in predicting the occurrence of AKD patients was compared using Delong's method. The discrimination was validated externally by the AUROC in the MIMIC III database.

We performed all statistical analyzes using R software version 4.0.5 (R Foundation for Statistical Computing).

RESULTS

Participants

As shown in **Figure 1**, a total of 5,629 patients were screened during the study period in the BFH. The initial research

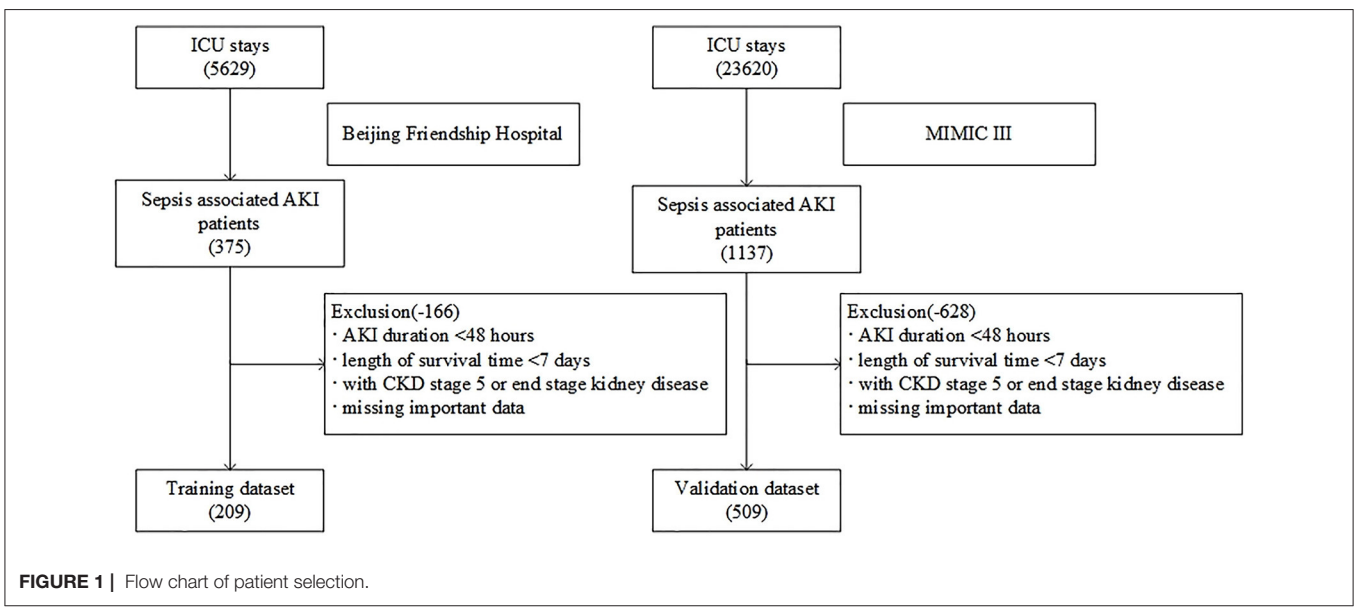


FIGURE 1 | Flow chart of patient selection.

TABLE 1 | Baseline characteristics of the Beijing Friendship Hospital (BFH) and Medical Information Mart for Intensive Care III (MIMIC III) cohorts.

	BFH cohort			MIMIC III cohort		
	Non-AKD (*n* = 93)	**AKD (*n* = 116)**	***P*-value**	**Non-AKD (*n* = 273)**	**AKD (*n* = 236)**	***P*-value**
Age, mean (SD)	54.7 (20.7)	64.5 (14.7)	<0.001	64.3 (16.5)	62.6 (18.1)	0.252
Male, (%)	58 (62.4)	74 (63.8)	0.832	160 (58.6)	117 (49.6)	0.041
BMI, kg/m^2, median [Q1, Q3]	26.6 (22.5, 30.4)	26.3 (22.4, 29.4)	0.837	27.3 (23.5, 32.2)	27.3 (23.3, 31.2)	0.782
Heart failure, *n* (%)	19 (20.4)	30 (25.9)	0.357	66 (24.2)	50 (21.2)	0.423
Hypertension, *n* (%)	45 (48.4)	74 (63.8)	0.025	21 (7.7)	30 (12.7)	0.060
Chronic obstructive pulmonary disease, *n* (%)	12 (12.9)	19 (16.4)	0.482	62 (22.7)	61 (25.8)	0.410
Chronic liver disease, *n* (%)	3 (3.2)	6 (5.2)	0.491	30 (11.0)	27 (11.4)	0.872
Diabetes mellitus, *n* (%)	35 (37.6)	66 (56.9)	0.006	76 (27.8)	52 (22.0)	0.132
Chronic kidney disease, *n* (%)	37 (39.8)	26 (22.4)	0.007	36 (13.2)	18 (7.6)	0.042
Charlson score, median [Q1, Q3]	2 (1, 3)	2 (1, 4)	0.001	2 (1, 3)	2 (1, 3)	0.729
Emergency department, *n* (%)	59 (63.4)	75 (64.7)	0.856	33 (12.1)	36 (15.3)	0.298
Surgery, *n* (%)	25 (26.9)	20 (17.2)	0.092	99 (36.3)	56 (23.7)	0.002
APS III, median [Q1, Q3]	45 (32, 62)	44.5 (32, 63)	0.779	40 (29, 55)	36.5 (26, 48)	0.035
SAPS II, median [Q1, Q3]	35 (25, 46)	35 (27, 44)	0.779	31 (23, 44)	31.5 (23, 41)	0.461
Non-renal SOFA at day 1, median [Q1, Q3]	3 (1, 6)	3 (1, 6)	0.310	3 (1, 5)	2 (1, 4)	0.044
Non-renal SOFA at day 3, median [Q1, Q3]	3 (1, 6)	3 (1, 6)	0.375	2 (1, 4)	2 (1, 4)	0.226
Delta non-renal SOFA, median [Q1, Q3]	0 (0, 0)	0 (0, 1)	<0.001	0 (0, 0)	0 (0, 0)	0.478
AKI stage, *n* (%)			<0.001			0.008
1	13 (14.0)	3 (2.6)		132 (48.4)	145 (61.4)	
2	33 (35.5)	20 (17.2)		91 (33.3)	55 (23.3)	
3	47 (50.5)	93 (80.2)		50 (18.3)	36 (15.3)	
Baseline creatinine, mg/dl, median [Q1, Q3]	0.7 (0.5,1.0)	0.8 (0.5, 1.1)	0.880	0.6 (0.5, 0.9)	0.60 (0.4, 0.9)	0.700
Creatinine at day 1, mg/dl, median [Q1, Q3]	1.3 (0.9, 2.2)	1.4 (0.9, 2.5)	0.857	1.1 (0.9, 1.5)	1.0 (0.8, 1.4)	0.050
Creatinine at day 3, mg/dl, median [Q1, Q3]	1.1 (0.9, 1.6)	1.2 (0.8, 2.0)	0.014	1.0 (0.8, 1.3)	0.9 (0.7, 1.4)	0.159
Delta creatinine, mg/dl, median [Q1, Q3]	−0.1 (−0.7, 0.0)	−0.1 (−0.40, 0.0)	0.001	−0.1 (−0.20, 0.0)	0.0 (−0.1, 0.0)	<0.001
Urine output at day1, ml/kg/h, median [Q1, Q3]	0.9 (0.4, 2.9)	0.9 (0.4, 2.8)	0.457	1.9 (0.7, 3.9)	1.9 (0.8, 4.0)	0.324
Urine output at day3, ml/kg/h, median [Q1, Q3]	1.1 (0.7, 1.7)	0.9 (0.3, 1.5)	<0.001	1.0 (0.6, 1.6)	1.1 (0.7, 2.1)	0.438
Delta urine output, ml/kg/h, median [Q1, Q3]	1.1 (0.7, 1.7)	0.9 (0.3, 1.5)	<0.001	−0.3 (−1.7, 0.0)	−0.2 (−1.0, 0.0)	0.219
Diuretics, *n* (%)	22 (23.7)	94 (81.0)	<0.001	126 (46.2)	91 (38.6)	0.084
Mechanical ventilation, *n* (%)	61 (65.6)	74 (63.8)	0.787	137 (50.2)	103 (43.6)	0.141
Renal toxic drugs, *n* (%)	46 (49.5)	89 (76.7)	<0.001	115 (42.1)	84 (35.6)	0.132

MIMIC III, Medical Information Mart for Intensive Care III; AKD, acute kidney disease; BMI, body mass index; APS III, Acute Physiological Score III; SPAS II, Simplified Acute Physiology Score II; SOFA, Sequential Organ Failure Assessment.

identified 23,620 ICU admissions from the MIMIC III database. In addition, 209 and 509 patients were assigned to the training dataset and validation dataset, respectively. Twenty-eight predictors were extracted from the database and included in the model. The occurrence of AKD rate was 55.5% (116 patients with AKD) in the training dataset and 46.4% (236 patients with AKD) in the validation dataset. A comparison of baseline characteristics between the AKD group and non-AKD group in BFH and MIMIC-III cohorts are recorded in **Table 1**. AKD patients were older and had higher Charlson score and delta non-renal SOFA; higher creatinine at day 3 and AKI stage; more medical history of hypertension, diabetes mellitus, and CKD; more application of diuretics and renal toxic drugs in the training dataset ($p < 0.05$), while they had a lower delta creatinine, urine output at day 3, and delta urine output ($p < 0.05$). Furthermore, comorbidities of CKD, higher AKI stage, and lower delta creatinine also showed similar results between AKD patients and non-AKD patients in the validation dataset ($p < 0.05$). Our study was reported according to the guidelines of the TRIPOD statement.

Model Development

In RNN-LSTM, as the validation loss was decreasing over time, the accuracy of the model increased (**Figure 2**). The LSTM has been trained up to 200 epochs to obtain the smallest loss and the greatest accuracy. Throughout the training process of 200 epochs, our training loss and validation loss had decreased and accuracy increased gradually, respectively. At the 200th epoch, the training loss and the validation loss are approximately the lowest, where the training accuracy and the validation accuracy reach 97.96 and 97.66%, respectively. We found that the training graph and the validation graph are quite similar. Thus, it can be

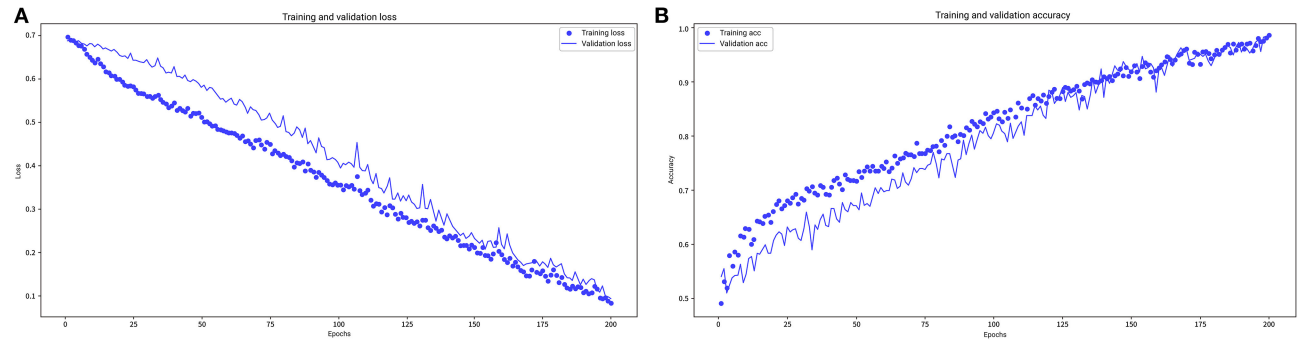


FIGURE 2 | Loss **(A)** and accuracy **(B)** vs. epoch graph (up to 200 epochs).

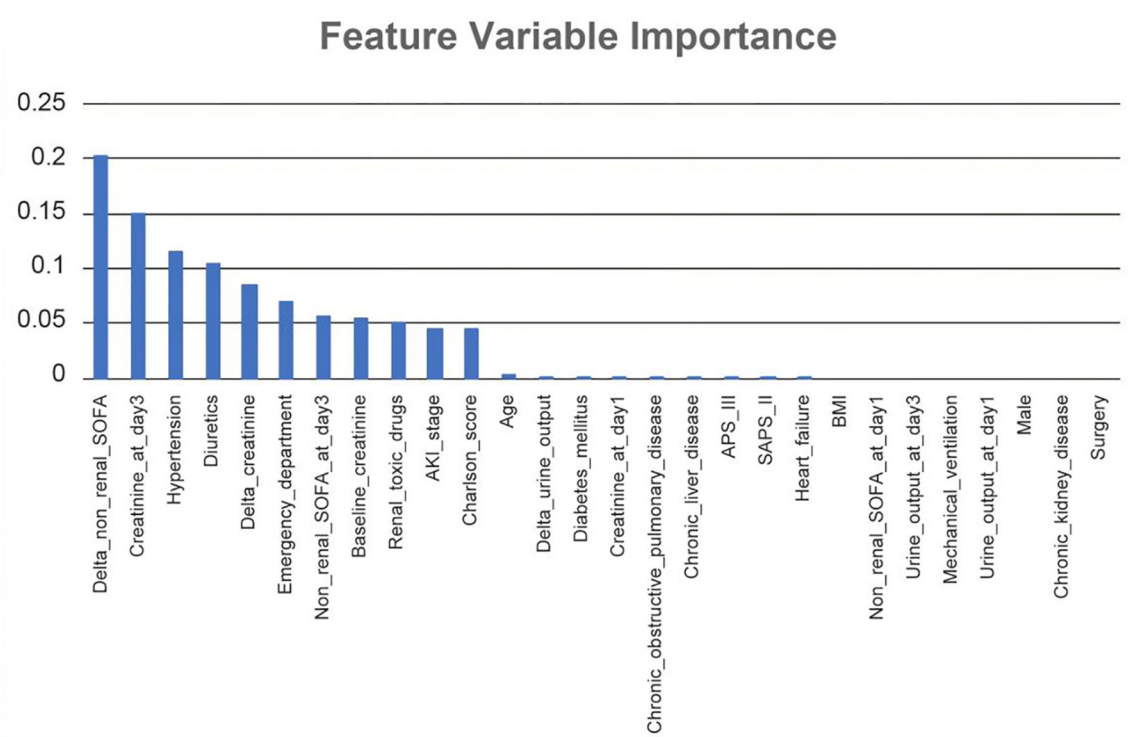


FIGURE 3 | Significance of the predictors in the Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM) model. All 28 important features regarding the development of the final predictive model are depicted.

concluded that the model is quite accurate. It is neither overfitting nor underfitting. The significance of the predictors in the RNN-LSTM model is presented in **Figure 3**. The feature variable importance showed that Δnon-renal SOFA had an important role. Other variables, such as creatinine on day 3, hypertension, and diuretics, also showed marked effects. As the decision trees algorithm has nodes that represent variables and conjunction that connects the nodes, the performance of this algorithm mainly depends on the number of nodes and tree size (31). We explored different ways to find the optimal performance of the decision trees algorithm by adjusting the number of nodes (**Figure 4**). We found that the optimal number of nodes that could minimize the decision trees' misclassification error rate was 10, where the complexity parameter was 0.018. Using this number of nodes, the decision trees' structure was pruned. Among these variables, Δnon-renal SOFA had a crucial role in the prediction of the occurrence of AKD. If Δnon-renal SOFA < 1, delta creatinine played an important role in the next decision. If Δnon-renal SOFA > 1, whether used diuretics or not was important. If Δnon-renal SOFA > 1 and patients did not receive diuretics, he/she was more likely to be diagnosed with AKD soon (**Figure 5**).

In logistic regression, twenty-eight variables were included in the LASSO regression analysis and narrowed down to 10 features in the LASSO regression model (**Figure 6**). Next, a model

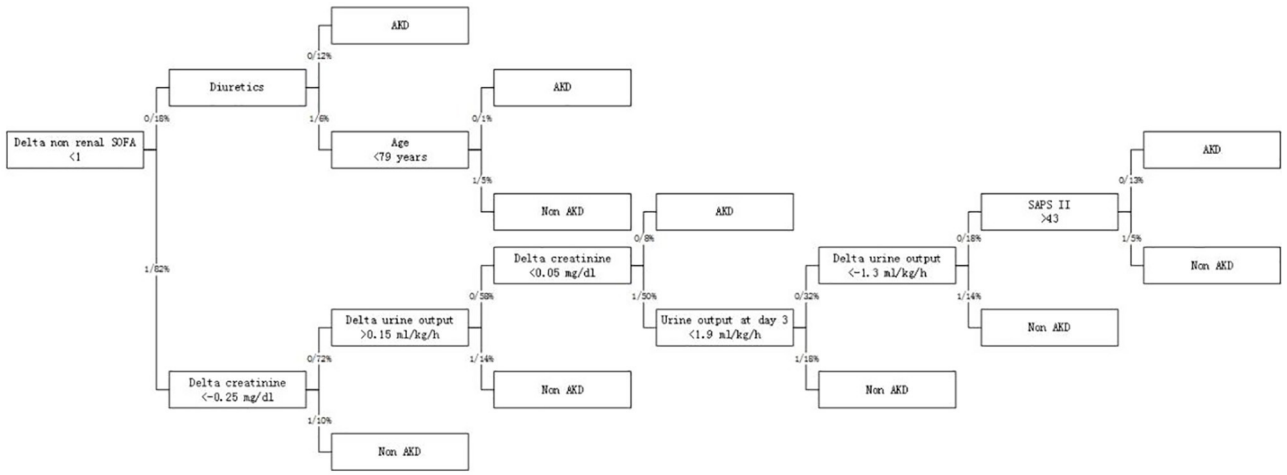


FIGURE 4 | Contribution of 28 variables in predicting the occurrence of patients with sepsis-associated AKD.

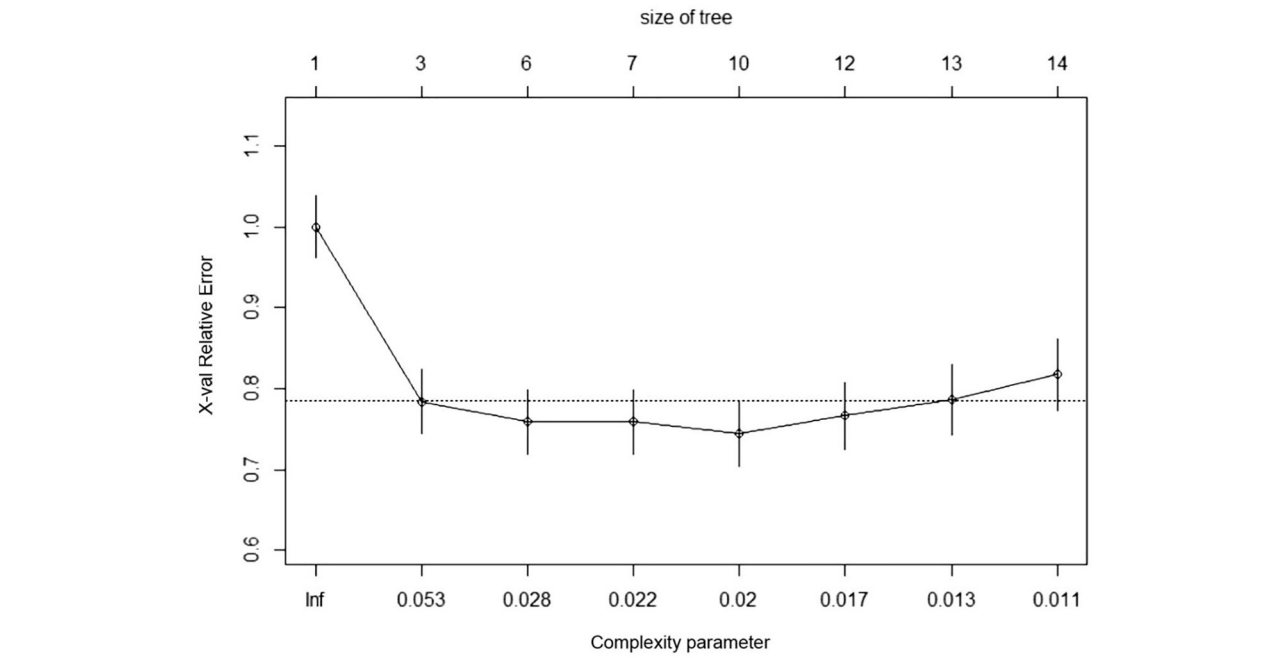


FIGURE 5 | Optimized decision tree for the classification of acute kidney disease (AKD)/non-AKD of patients.

integrating age, combined with hypertension, diabetes mellitus, CKD, delta non-renal SOFA, AKI stage, delta creatinine, delta urine output, diuretics, and nephrotoxic drugs was established using the training dataset. Based on this model, a nomogram was plotted to predict the probability of the occurrence of AKD patients (**Figure 7**). The calibration curve was described using the bootstrap method for both, the training and validation datasets (**Figure 8A**). The apparent line and a bias-corrected line only slightly deviated from the ideal line, indicating a good agreement between the prediction and reality. The DCA curve was plotted to perform a clinical application of this nomogram. In the training dataset, clinical intervention guided by this nomogram provided a greater net benefit when the threshold probability was within 0.01 and 0.71 (**Figure 8B**).

Model Performance

In the training dataset, we evaluated the discrimination of three models. RNN-LSTM was well-discriminated in the external validation dataset (AUROC: 1), which was greater than decision trees and logistic regression (AUROC: decision

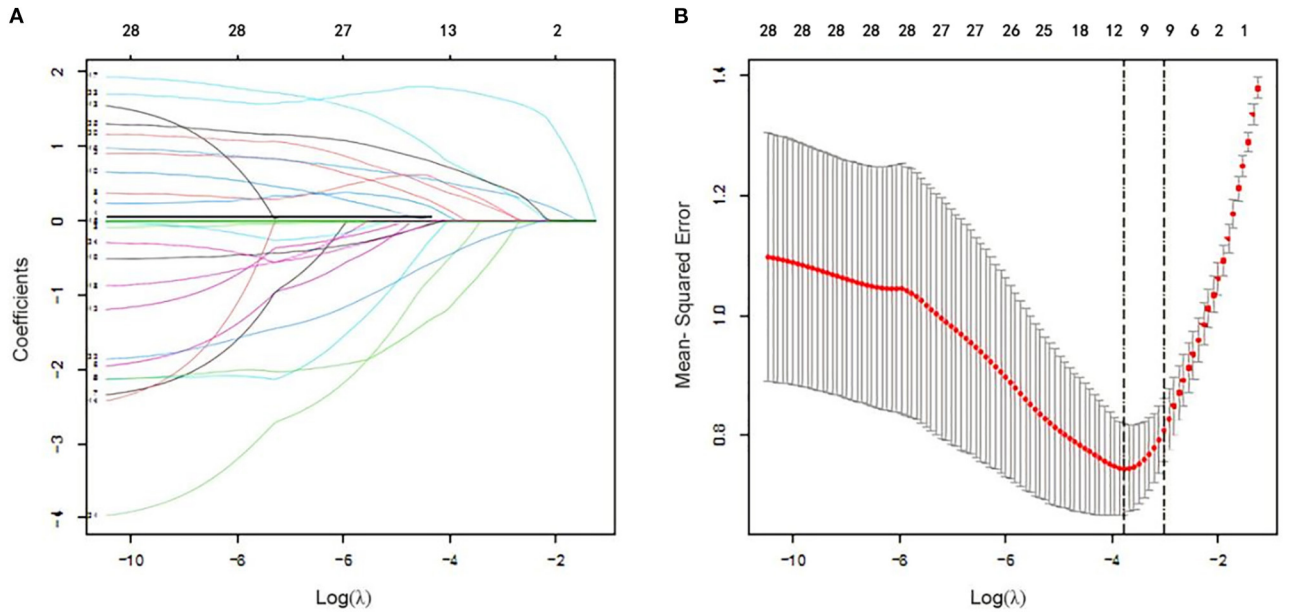


FIGURE 6 | Clinical feature selection using the Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression. **(A)** Optimal parameter (lambda) selection in the LASSO logistic regression. The black vertical lines were drawn at the optimal values by using the minimum criteria and the one SE of the minimum criteria (the 1-SE criteria). **(B)** LASSO coefficient profiles of the 28 features. A coefficient profile plot was produced against the log (lambda) sequence.

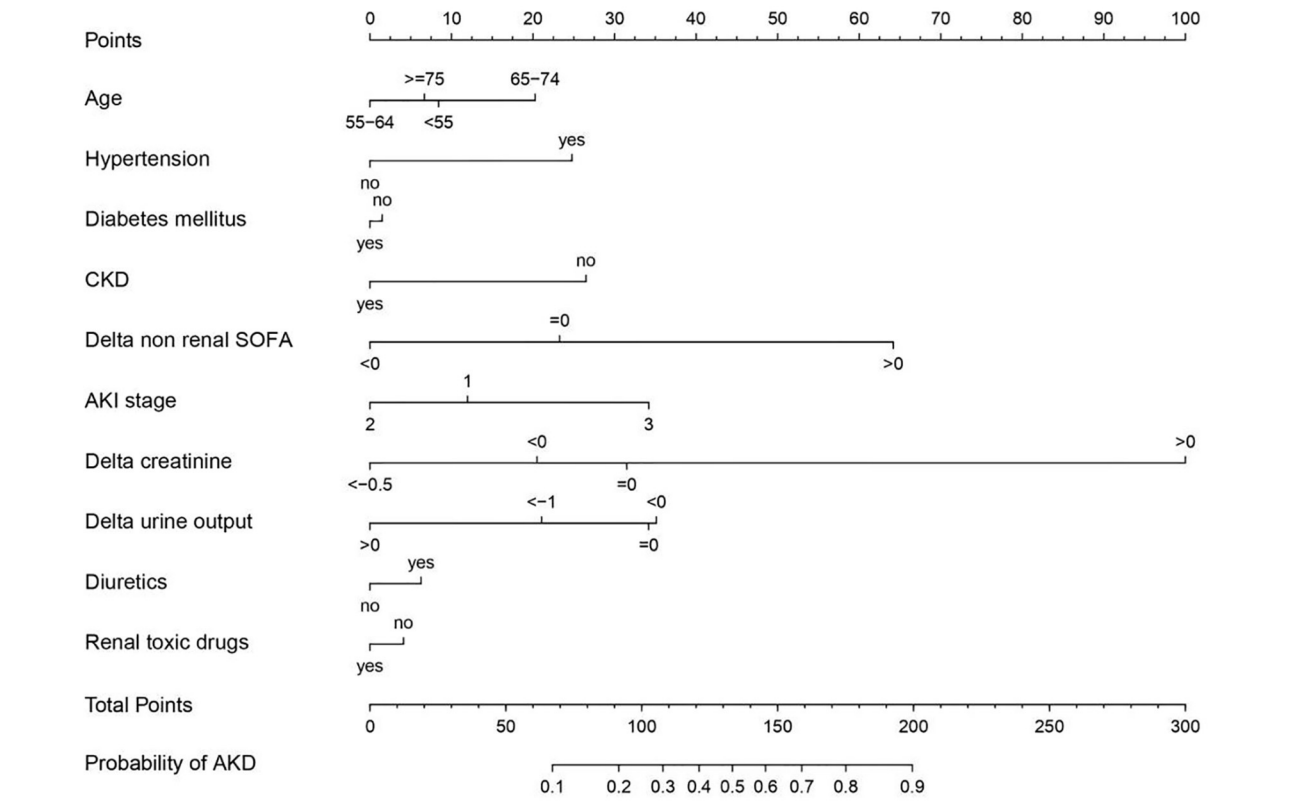


FIGURE 7 | Nomogram developed based on the training dataset with the incorporation of age, combined with hypertension, diabetes mellitus, chronic kidney disease (CKD), delta non-renal Sequential Organ Failure Assessment (SOFA), acute kidney injury (AKI) stage, delta creatinine, delta urine output, diuretics, and renal toxic drugs.

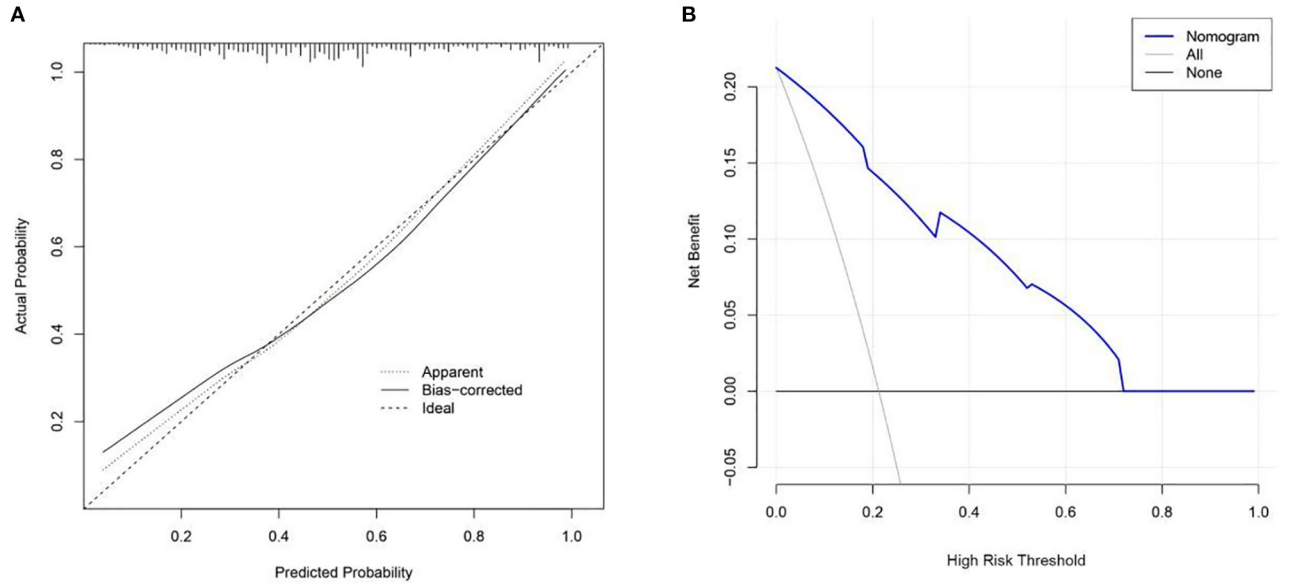


FIGURE 8 | Calibration curves **(A)** and decision curve analysis **(B)** for nomogram.

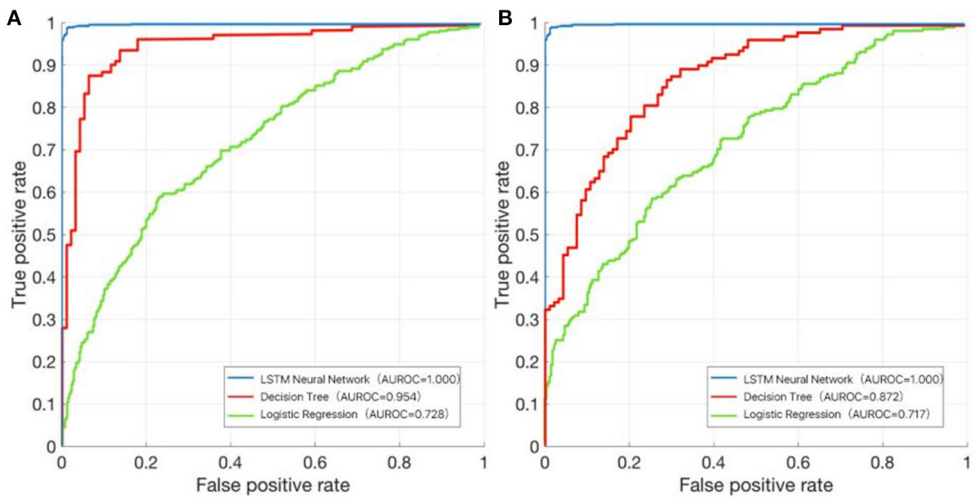


FIGURE 9 | The area under the receiver operating characteristic (AUROC) curve of the RNN-LSTM, decision trees, and logistic regression. **(A)** Training dataset; **(B)** Validation dataset.

trees 0.954, logistic regression 0.728; **Figure 9A**). In the validation dataset, among RNN-LSTM, decision trees, and logistic regression algorithms, the RNN-LSTM algorithm showed the highest performance with an AUROC of 1.000, followed by the decision trees with an AUROC of 0.872. Logistic regression had the least predictive accuracy, with an AUROC of 0.717. All machine learning models, except the logistic regression model, showed good discrimination ability in the training and validation datasets. In the training and validation datasets, the RNN-LSTM algorithm achieved the best performance among the four models (**Figure 9B**).

DISCUSSION

In the present study, a total of 209 patients from BFH were included, with 55.5% of them diagnosed as having AKD. Using the data from BFH and MIMIC III records, we successfully developed and validated machine learning models to predict the occurrence of AKD in patients with AKI.

Since the diagnostic criteria for AKD were released in ADQI-16, several investigations have been undertaken on the epidemiology of AKD. Kellum et al. reported the incidence rate of AKD as 36.2% in ICU patients (4). Federspiel et

al. showed the incidence rate of sepsis-associated AKD as 32.4% in critically ill patients (5). Peerapornratana et al. reported the incidence rate of sepsis-associated AKD in patients dying within 7 days was 33.6% (161/479) from the first day of being diagnosed with AKI (11). Our studies showed an AKD diagnosis rate of 55.5%. This higher rate could be attributed to the exclusion of patients with an AKI duration of <3 days.

Ostermann et al. suggested that nephrotoxic drugs increase the risk of renal function impairment (32). Drugs are among the main causes of AKI. Its pathogenesis included acute tubular necrosis, tubular obstruction by crystals or casts, and interstitial nephritis induced by drugs and their metabolites (33). Our study shows that nephrotoxic drugs increase the incidence of AKD, possibly because they deteriorate renal function.

There has been a controversy about whether the application of diuretics can improve renal function in recent years. A Phase II Randomized Blinded Controlled Trial of the Effect of furoSemide in Critically Ill Patients With eARly Acute Kidney Injury (SPARK-RCT) study showed that diuretics improved neither the recovery rate of AKI nor the prognosis of the patients (34). The study of Zhao et al. reported that administering diuretics improved renal function in patients on the MIMIC III database (35). Our research shows that the use of diuretics may be related to the low incidence of AKD. The effective use of diuretics can reflect the recovery of the patients' renal function, but it may not change it. More research is needed to further clarify the role of diuretics in improving renal function.

There are some studies on the prediction of AKD in hospitalized patients with AKI. Zhao et al. used multivariable logistic regression analysis with the LASSO method to select features and build a nomogram (36). The model displayed good predictive power with an AUROC of 0.834 (95% CI:0.773–0.895) in the training dataset and an AUROC of 0.851 (95% CI:0.753–0949) in the validation dataset. Yan et al. also established a prediction model using multivariable logistic regression analysis (37). The 8-variable model showed good discrimination and calibration in predicting AKD stage 2–3 with the AUROC being 0.85 (95% CI:0.83–0.87). Xiao et al. established a prediction model using multivariable logistic regression analysis. This model showed a large AUROC (0.879 ± 0.009, 0.879 ± 0.011) and had stable sensitivity (81 and 82%) and specificity (81 and 80%) in derivation cohort and validation dataset, respectively (38). In our study, the AUROCs of the logistic regression model were 0.728 (training dataset) and 0.717 (validation dataset), which were lower than the above studies. This may be due to differences in the study population. A study by Tuan et al. studied sepsis-associated AKI patients, however, they predicted progression to chronic kidney disease rather than AKD (39). Therefore, to our knowledge, this is the first study to use longitudinal data to predict the occurrence of AKD with the application of machine learning.

To identify AKD patients, an important strength of our study was the use of new criteria of sepsis-associated AKI, and this method would overcome some inherent weaknesses of using hospital discharge data (40, 41). The delta non-renal SOFA contains only 5 simple variables recorded in clinical routines. Therefore, if implemented, the delta non-renal SOFA will not require manual input of additional variables as the model is based on variables routinely collected. In our study, for predicting the occurrence of AKD, the delta non-renal SOFA score had high discriminatory power. The delta non-renal SOFA is simple for calculation and easy to use and has robust discrimination and calibration. To predict the occurrence of AKD patients with sepsis, ICU physicians could use the delta non-renal SOFA and improve clinical decision-making at the bedside. Moreover, the predictor variables that we used were quite universally obtained in the emergency department. After further validation and recalibration, the delta non-renal SOFA appeared to have the potential to help emergency department clinicians triage decisions and ICU placement.

Limitations

The study has the following limitations. First, we chose to analyze the patients admitted to the ICU with sepsis. There were certainly patients who had been diagnosed with sepsis before or after the ICU admission, but we limited our study population to those who fulfilled sepsis-3 criteria during their 1st day in ICU. Second, we have a limited number of patients and a small sample size, but we conducted an external validation by using the data of 509 sepsis-associated AKI patients from the MIMIC III database, and the results indicated that the calibration of delta non-renal SOFA was relatively well with accordance of occurrence of AKD. Finally, we prepared our dataset from the retrospective database, and the outcomes of sepsis-associated AKI patients could have changed over time due to the update of treatment guidelines and advances in treatment and diagnostic technology.

CONCLUSION

Machine learning could be applied to the predictive AKD, and it is where the RNN-LSTM model works the best. The non-renal SOFA plays an important role in predicting the AKD.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by institutional review boards of Beijing Friendship Hospital and the Massachudatasetts Institute of Technology. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JH and MD conceived the idea, performed the analysis, and drafted the manuscript. JH and JL interpreted the

results and helped to revise the manuscript. JL and MD helped to frame the idea of the study and helped to analyze the data. All authors read and approved the final manuscript.

FUNDING

This work was supported in part by grants from the Beijing Key Clinical Specialty Excellence Project.

REFERENCES

1. Hoste EA, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive Care Med.* (2015) 41:1411–23. doi: 10.1007/s00134-015-3934-7
2. Bouchard J, Acharya A, Cerda J, Maccariello ER, Madarasu RC, Tolwani AJ, et al. A prospective international multicenter study of AKI in the intensive care unit. *Clin J Am Soc Nephrol.* (2015) 10:1324–31. doi: 10.2215/CJN.04360514
3. See EJ, Jayasinghe K, Glassford N, Bailey M, Johnson DW, Polkinghorne KR, et al. Long-term risk of adverse outcomes after acute kidney injury: a systematic review and meta-analysis of cohort studies using consensus definitions of exposure. *Kidney Int.* (2019) 95:160–72. doi: 10.1016/j.kint.2018.08.036
4. Kellum JA, Sileanu FE, Bihorac A, Hoste EA, Chawla LS. Recovery after acute kidney injury. *Am J Respir Crit Care Med.* (2017) 195:784–91. doi: 10.1164/rccm.201604-0799OC
5. Federspiel CK, Itenov TS, Mehta K, Hsu RK, Bestle MH, Liu KD. Duration of acute kidney injury in critically ill patients. *Ann Intensive Care.* (2018) 8:30. doi: 10.1186/s13613-018-0374-x
6. Mehta S, Chauhan K, Patel A, Patel S, Pinotti R, Nadkarni GN, et al. The prognostic importance of duration of AKI: a systematic review and meta-analysis. *BMC Nephrol.* (2018) 19:91. doi: 10.1186/s12882-018-0876-7
7. Chawla LS, Bellomo R, Bihorac A, Goldstein SL, Siew ED, Bagshaw SM, et al. Acute kidney disease and renal recovery: consensus report of the Acute Disease Quality Initiative (ADQI) 16 workgroup. *Nat Rev Nephrol.* (2017) 13:241–57. doi: 10.1038/nrneph.2017.2
8. Fujii T, Uchino S, Takinami M, Bellomo R. Subacute kidney injury in hospitalized patients. *Clin J Am Soc Nephrol.* (2014) 9:457–61. doi: 10.2215/CJN.04120413
9. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care.* (2019) 23:112. doi: 10.1186/s13054-019-2411-z
10. Zimmerman LP, Reyfman PA, Smith ADR, et al. Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements. *BMC Med Inform Decis Mak.* (2019) 19:16. doi: 10.1186/s12911-019-0733-z
11. Peerapornratana S, Priyanka P, Wang S, Smith A, Singbartl K, Palevsky PM, et al. Sepsis-associated acute kidney disease. *Kidney Int Rep.* (2020) 5:839–50. doi: 10.1016/j.ekir.2020.03.005
12. Hsu CK, Wu IW, Chen YT, Tsai TY, Tsai FC, Fang JT, et al. Acute kidney disease stage predicts outcome of patients on extracorporeal membrane oxygenation support. *PLoS ONE.* (2020) 15:e0231505. doi: 10.1371/journal.pone.0231505
13. Chen YT, Jenq CC, Hsu CK, Yu YC, Chang CH, Fan PC, et al. Acute kidney disease and acute kidney injury biomarkers in coronary care unit patients. *BMC Nephrol.* (2020) 21:207. doi: 10.1186/s12882-020-01872-z
14. Lee BJ, Hsu CY, Parikh R, McCulloch CE, Tan TC, Liu KD, et al. Predicting renal recovery after dialysis-requiring acute kidney injury. *Kidney Int Rep.* (2019) 4:571–81. doi: 10.1016/j.ekir.2019.01.015
15. Fiorentino M, Tohme FA, Murugan R, Kellum JA. Plasma biomarkers in predicting renal recovery from acute kidney injury in critically Ill patients. *Blood Purif.* (2019) 48:253–61. doi: 10.1159/000500423
16. Cho JS, Shim JK, Lee S, Song JW, Choi N, Lee S, et al. Chronic progression of cardiac surgery associated acute kidney injury: intermediary role of acute kidney disease. *J Thorac Cardiovasc Surg.* (2021) 161:681–8.e3. doi: 10.1016/j.jtcvs.2019.10.101
17. Hu P, Song L, Liang H, Chen Y, Wu Y, Zhang L, et al. Prospective model for predicting renal recovery in cardiac surgery patients with acute kidney injury requiring renal replacement therapy. *Nephrology.* (2021) 26:586–93. doi: 10.1111/nep.13878
18. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data.* (2016) 3:160035. doi: 10.1038/sdata.2016.35
19. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* (2015) 162:W1–73. doi: 10.7326/M14-0698
20. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA.* (2016) 315:801–10. doi: 10.1001/jama.2016.0287
21. Kidney Disease: Improving Global Outcomes (KDIGO) Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury. *Kidney Inter.* (2012) 2:1–138.
22. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney inter.* (2013) 3:1–150.
23. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res.* (2017) 26:796–808. doi: 10.1177/0962280214558972
24. Lipton ZC, Kale DC, Elkan C, Wetzel RJapa. Learning to diagnose with LSTM recurrent neural networks (2015).
25. Khalil K, Eldash O, Kumar A, Bayoumi M. Economic LSTM approach for recurrent neural networks. *IEEE Transac Circ Syst II: Expr Briefs.* (2019) 66:1885–89. doi: 10.1109/TCSII.2019.2924663
26. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. (2017). doi: 10.1201/9781315139470
27. Drummond C, Holte RC. Exploiting the cost (In) sensitivity of decision tree splitting criteria. *ICML.* (2000) 1:8.
28. Tibshirani R. Regression shrinkage and selection *via* the lasso. *J R Stat Soc: Ser B.* (1996) 58:267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
29. Zhang Z, Rousson V, Lee WC, Ferdynus C, Chen M, Qian X, et al. Decision curve analysis: a technical note. *Ann Transl Med.* (2018) 6:308. doi: 10.21037/atm.2018.07.02
30. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* (2006) 26:565–74. doi: 10.1177/0272989X06295361
31. Bradford JP, Kunz C, Kohavi R, Brunk C, Brodley CE. Pruning decision trees with misclassification costs. In: *Paper presented at: European Conference on Machine Learning.* (1998). doi: 10.1007/BFb0026682
32. Ostermann M, Zarbock A, Goldstein S, Kashani K, Macedo E, Murugan R, et al. Recommendations on acute kidney injury biomarkers from the acute disease quality initiative consensus conference: a consensus statement. *JAMA Netw Open.* (2020) 3:e2019209. doi: 10.1001/jamanetworkopen.2020.19209
33. Kwiatkowska E, Domański L, Dziedziejko V, Kajdy A, Stefańska K, Kwiatkowski S. The mechanism of drug nephrotoxicity and the methods for preventing kidney damage. *Int J Mol Sci.* (2021) 22:6109. doi: 10.3390/ijms22116109
34. Bagshaw SM, Gibney RTN, Kruger P, Hassan I, McAlister FA, Bellomo R. The effect of low-dose furosemide in critically ill patients with early acute kidney injury: A pilot randomized blinded controlled trial (the SPARK study). *J Crit Care.* (2017) 42:138–46. doi: 10.1016/j.jcrc.2017.07.030
35. Zhao GJ, Xu C, Ying JC, Lü WB, Hong GL Li MF, et al. Association between furosemide administration and outcomes in critically ill patients

with acute kidney injury. *Crit Care.* (2020) 24:75. doi: 10.1186/s13054-020-2798-6
36. Zhao H, Liang L, Pan S, Liu Z, Liang Y, Qiao Y, et al. Diabetes mellitus as a risk factor for progression from acute kidney injury to acute kidney disease: a specific prediction model. *Diabetes Metab Syndr Obes.* (2021) 14:2367–79. doi: 10.2147/DMSO.S307776
37. Yan P, Duan XJ, Liu Y, Wu X, Zhang NY, Yuan F, et al. Acute kidney disease in hospitalized acute kidney injury patients. *PeerJ.* (2021) 9:e11400. doi: 10.7717/peerj.11400
38. Xiao YQ, Cheng W, Wu X, Yan P, Feng LX, Zhang NY, et al. Novel risk models to predict acute kidney disease and its outcomes in a Chinese hospitalized population with acute kidney injury. *Sci Rep.* (2020) 10:15636. doi: 10.1038/s41598-020-72651-x
39. Tuan PNH, Quyen DBQ, Van Khoa H, Loc ND, Van My P, Dung NH, et al. Serum and urine neutrophil gelatinase-associated lipocalin levels measured at admission predict progression to chronic kidney disease in sepsis-associated acute kidney injury patients. *Dis Markers.* (2020) 2020:8883404. doi: 10.1155/2020/8883404
40. Ford DW, Goodwin AJ, Simpson AN, Johnson E, Nadig N, Simpson KN, et al. Severe sepsis mortality prediction model and score for use with administrative data. *Crit Care Med.* (2016) 44:319–27. doi: 10.1097/CCM.0000000000001392
41. Johnson AEW, Aboab J, Raffa JD, Pollard TJ, Deliberato RO, Celi LA, et al. A comparative analysis of sepsis identification methods in an electronic database. *Crit Care Med.* (2018) 46:494–9. doi: 10.1097/CCM.0000000000002965

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 10 December 2021
doi: 10.3389/fmed.2021.792689

A Machine Learning Approach for the Prediction of Traumatic Brain Injury Induced Coagulopathy

Fan Yang[1†], *Chi Peng*[2†], *Liwei Peng*[3†], *Jian Wang*[3], *Yuejun Li*[1] *and Weixin Li*[3*]

[1] *Department of Plastic Surgery and Burns, Tangdu Hospital, Fourth Military Medical University, Xi'an, China,* [2] *Department of Health Statistics, Second Military Medical University, Shanghai, China,* [3] *Department of Neurosurgery, Tangdu Hospital, Fourth Military Medical University, Xi'an, China*

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Luo Zhe,
Fudan University, China
Arulselvi Subramanian,
All India Institute of Medical Sciences, India
Gaurav Chhabra,
All India Institute of Medical Sciences, Bhubaneswar, India

***Correspondence:**
Weixin Li
tangdunaowai@163.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *11 October 2021*
Accepted: *08 November 2021*
Published: *10 December 2021*

Citation:
Yang F, Peng C, Peng L, Wang J, Li Y and Li W (2021) A Machine Learning Approach for the Prediction of Traumatic Brain Injury Induced Coagulopathy. Front. Med. 8:792689. doi: 10.3389/fmed.2021.792689

Background: Traumatic brain injury-induced coagulopathy (TBI-IC), is a disease with poor prognosis and increased mortality rate.

Objectives: Our study aimed to identify predictors as well as develop machine learning (ML) models to predict the risk of coagulopathy in this population.

Methods: ML models were developed and validated based on two public databases named Medical Information Mart for Intensive Care (MIMIC)-IV and the eICU Collaborative Research Database (eICU-CRD). Candidate predictors, including demographics, family history, comorbidities, vital signs, laboratory findings, injury type, therapy strategy and scoring system were included. Models were compared on area under the curve (AUC), accuracy, sensitivity, specificity, positive and negative predictive values, and decision curve analysis (DCA) curve.

Results: Of 999 patients in MIMIC-IV included in the final cohort, a total of 493 (49.35%) patients developed coagulopathy following TBI. Recursive feature elimination (RFE) selected 15 variables, including international normalized ratio (INR), prothrombin time (PT), sepsis related organ failure assessment (SOFA), activated partial thromboplastin time (APTT), platelet (PLT), hematocrit (HCT), red blood cell (RBC), hemoglobin (HGB), blood urea nitrogen (BUN), red blood cell volume distribution width (RDW), creatinine (CRE), congestive heart failure, myocardial infarction, sodium, and blood transfusion. The external validation in eICU-CRD demonstrated that adapting boosting (Ada) model had the highest AUC of 0.924 (95% CI: 0.902–0.943). Furthermore, in the DCA curve, the Ada model and the extreme Gradient Boosting (XGB) model had relatively higher net benefits (ie, the correct classification of coagulopathy considering a trade-off between false- negatives and false-positives)—over other models across a range of threshold probability values.

Conclusions: The ML models, as indicated by our study, can be used to predict the incidence of TBI-IC in the intensive care unit (ICU).

Keywords: traumatic brain injury-induced coagulopathy, TBI-IC, machine learning, external validation, model interpretation

INTRODUCTION

Traumatic brain injury (TBI) is still one of the leading causes of death and disability worldwide with over 10 million people hospitalized every year (1). It is common to witness the alterations of the coagulative system and disturbed coagulation function in TBI patients. Results from previous studies indicated that two in three patients with severe TBI manifested coagulation system abnormalities upon admission to the emergency department, and then continued to worsen (2, 3). And the overall mortality of TBI-induced coagulopathy (TBI-IC) attains 17–86% (4–6). TBI-IC is characterized by both hypo-coagulopathy with prolonged bleeding or hyper-coagulopathy with an increased prothrombotic tendency, or both (4, 7). Previous study unearthed that coagulopathy following TBI was related to higher mortality and prolonged intensive care unit (ICU) stay (8). In early stage, potential mechanisms include the dysfunction of the coagulation cascade and hyperfibrinolysis, both of which contribute to hemorrhagic progression. Later, a poorly defined prothrombotic stage emerges, partly caused by fibrinolysis shutdown and hyperactive platelets (9–11).

Undoubtedly, it is imperative to promote the early identification of TBI-IC in a timely way. Laboratory assays, including international normalized ratio (INR) and thromboelastogram are widely used to diagnose TBI-IC. Nonetheless, these assays have limited value in predicting coagulopathy before it develops. In recent years, as a field of artificial intelligence, machine learning (ML) is able to learn from data based on computational modeling. Likewise, ML can fit high-order relationships between covariates and outcomes in data-rich environments (12–14).

This study aimed to determine whether ML algorithms using demographic, comorbidities, laboratory examinations and other variables could predict TBI-IC with considerable accuracy and identify factors contributing to the prediction power.

MATERIALS AND METHODS

Data Source

We conducted this retrospective study based on two sizeable critical care databases, the Medical Information Mart for Intensive Care (MIMIC)-IV version 1.0 (15) and eICU Collaborative Research Database (eICU-CRD) version 1.2 (16). In brief, the MIMIC-IV database, an updated version of MIMIC-III, incorporated comprehensive, de-identified data of patients admitted to the ICUs at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2008 and 2019, containing data from 383220 distinct admissions (single center). The other database, eICU-CRD, was a multicenter, freely available, sizeable database with de-identified high granularity health data associated for over 200,000 admissions to ICUs across the United States between 2014 and 2015. This study was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA). Requirement for individual patient consent was waived because the study did not impact clinical care and all protected health information was deidentified. One author (CP) has obtained access to both databases and was responsible for data extraction (Certification number: 41657645). The study was reported in accordance to the REporting of studies Conducted using Observational Routinely collected health Data (RECORD) statement (17).

Participant Selection

Inclusion criteria were patients with moderate and severe TBI [msTBI: defined as Glasgow Coma Score (GCS) =< 12]. People with an age of less than 16 years old, ICU stays less than 48 h, and no coagulation index within 24 h of ICU admission were excluded from the study. Moreover, for patients with ICU admissions more than once, only data of the first ICU admission of the first hospitalization were included in the analysis.

Predictors of Coagulopathy

A total of 53 predictor variables for the ML models were initially included. Specifically, in this study, the data were extracted from MIMIC-IV and eICU-CRD including age, gender, race, family history of stroke. Coexisting disorders were also collected based on the recorded International Classification of Diseases (ICD)-9 and ICD-10 codes. Then, the Charlson comorbidity index (CCI) was calculated from its component variables [myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, chronic pulmonary disease, rheumatic disease, peptic ulcer disease, diabetes, paraplegia, renal disease, malignant cancer, severe liver disease, metastatic solid tumor and acquired immunodeficiency syndrome (AIDS)]. Lastly, we extracted data containing vital signs, laboratory findings, injury type, different therapy strategies and scoring system on the first day of ICU admission. Details of missing data can be seen in **Supplementary Table 1**.

Outcome

In accordance to previous literature, the following parameters were considered for diagnosing coagulopathy: an activated partial thromboplastin time (APTT) $>$ 40s, an INR $>$ 1.4, or platelet (PLT) counts $< 100 \times 10^9$/L (4, 18).

Statistical Analysis

Values were presented as the means with standard deviations (if normal) or medians with interquartile ranges (IQR) (if non-normal) for continuous variables, and total numbers with percentages for categorical variables. Proportions were compared using χ^2 test or Fisher exact tests while continuous variables were compared using the *t* test or Wilcoxon rank sum test, as appropriate.

In this study, recursive feature elimination (RFE) as a feature selection method was used to select the most relevant features. In short, RFE recursively fits a model based on smaller feature sets until a specified termination criterion is reached. In each loop, in the trained model, features were ranked based on their importance. Finally, dependency and collinearity were eliminated. Features were then considered in groups of 15/25/35/45/ALL (ALL = 53 variables, as represented in **Figure 1**) organized by the ranks obtained after the feature

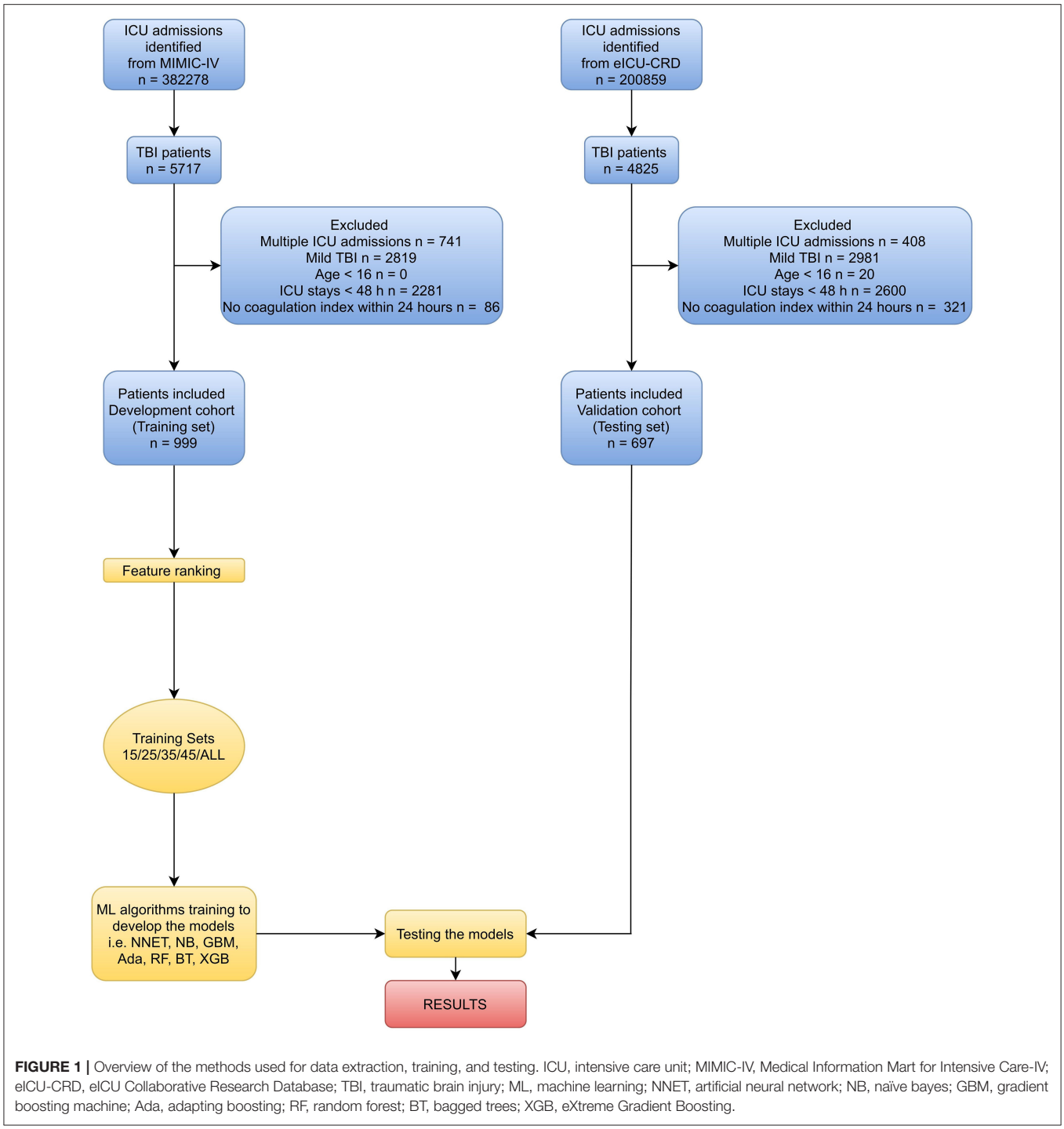


FIGURE 1 | Overview of the methods used for data extraction, training, and testing. ICU, intensive care unit; MIMIC-IV, Medical Information Mart for Intensive Care-IV; eICU-CRD, eICU Collaborative Research Database; TBI, traumatic brain injury; ML, machine learning; NNET, artificial neural network; NB, naïve bayes; GBM, gradient boosting machine; Ada, adapting boosting; RF, random forest; BT, bagged trees; XGB, eXtreme Gradient Boosting.

selection method. To find the optimal hyperparameters, 10-fold cross-validation was used as a resampling method. In each iteration, every nine folds were used as training subset, and the remaining one fold was processed to tune the hyperparameters. This training-testing process was repeated thirty times. And in this way, each sample would be involved in the training model, and also participated in the testing model, so that all data were used to the greatest extent.

In this study, we employed seven diverse ML algorithms to develop models, containing artificial neural network (NNET), naïve bayes (NB), gradient boosting machine (GBM), adapting boosting (Ada), random forest (RF), bagged trees (BT), and eXtreme Gradient Boosting (XGB). Initially, we conducted internal validation on the development sets to quantify optimism in the predictive performance and evaluate stability of the prediction model. Bootstrap Resampling technique with 100

iterations was used to evaluate the internal validity of each model. External validation of the models was performed in eICU-CRD. All the models were assessed in multiple dimensions regarding their model performance. The median and 95% confidence intervals of area under the curve (AUC) were calculated, where an AUC value of 1.0 means perfect discrimination and 0.5 represents no discrimination. And the accuracy, sensitivity, specificity, negative predictive value, and positive predictive value were also calculated. Additionally, to determine the clinical usefulness of the included variables by quantifying the net benefit at different threshold probabilities, we conducted the decision curve analysis (DCA) (19). Finally, the "Shiny" package in the R was used to construct a visual data analysis platform.

All analyses were performed by the statistical software packages R version 4.0.2 (http://www.R-project.org, The R Foundation). In our study, we used the "Caret" R packages to achieve the process. *P* values less than 0.05 (two-sided test) were considered as statistically significant.

RESULTS

Baseline Characteristics

Variable values on the first day of the TBI patients in MIMIC-IV were analyzed. As shown in **Figure 1** and **Table 1** of 5717 TBI patients in MIMIC-IV, 999 were included in the final cohort. A total of 493 patients developed coagulopathy, whereas 506 patients did not. A cohort of 285 patients with coagulopathy following TBI in eICU-CRD was included as external dataset (**Supplementary Table 2**). The process of data extraction, training preparation, data testing via different ML algorithms is depicted in **Figure 1**. People who had coagulopathy were more likely to be female, with family history of stroke, myocardial infarction, congestive heart failure, peripheral vascular disease, cerebrovascular disease, renal disease, malignant cancer, severe liver disease, metastatic solid tumor as well as having higher CCI, heart rate, respiratory rate, red blood cell volume distribution width (RDW), INR, lactate, buffer excess (BE), FiO_2, chloride, sodium, glucose, creatinine (CRE), blood urea nitrogen (BUN), blood transfusion, sepsis related organ failure assessment (SOFA), acute physiology score III (APSIII), and longer APTT, prothrombin time (PT), mechanical ventilation (MV). Furthermore, they were less likely to have dementia, cerebral contusion, with lower temperature, mean artery pressure (MAP), red blood cell (RBC), white blood cell (WBC), hemoglobin (HGB), PLT, hematocrit (HCT), pH, bicarbonate, PaO_2/FiO_2, calcium, urine output, and GCS.

Variable Importance

A total of 15 important predictors (**Figure 2**) was selected by the RFE algorithm, including INR, PT, SOFA, APTT, PLT, HCT, RBC, HGB, BUN, RDW, CRE, congestive heart failure, myocardial infarction, sodium, and blood transfusion. Then, these 15 variables were used in all the subsequent analysis for all models in both training and testing sets.

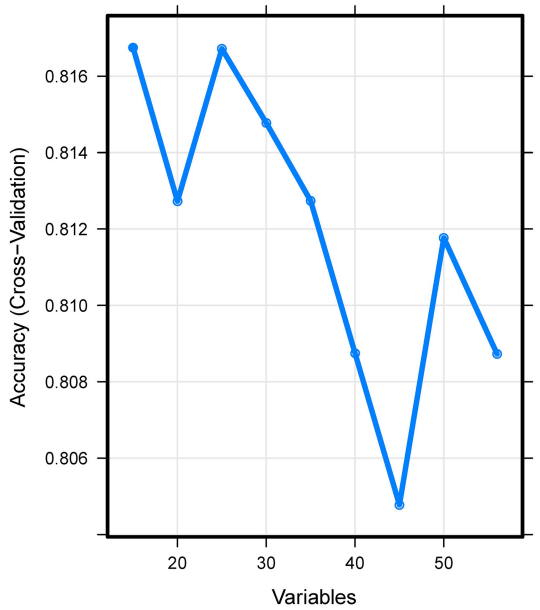


FIGURE 2 | Association between the number of variables allowed to be considered at each split and the prediction accuracy in the REF algorithm. REF, recursive feature elimination.

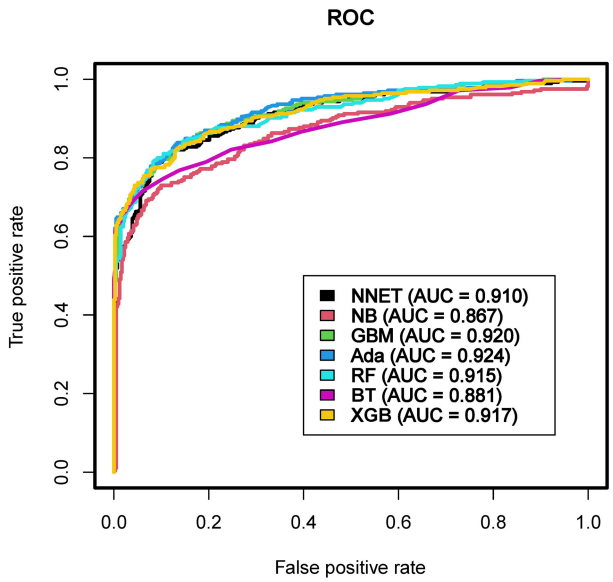


FIGURE 3 | Area under the curve of receiver operating characteristic curve by machine learning models in the validation cohort. ROC, receiver operate characteristics; AUC, area under the curve; NNET, artificial neural network; NB, naïve bayes; GBM, gradient boosting machine; Ada, adapting boosting; RF, random forest; BT, bagged trees; XGB, eXtreme Gradient Boosting.

Prediction Performance in eICU-CRD

The discriminatory abilities of all models for the prediction of coagulopathy are in **Figure 3** and **Table 2**. Within the training set, the NNET, NB, GBM, Ada, RF, BT and XGB models were

TABLE 1 | Baseline characteristics of the MIMIC-IV cohorts.

	MIMIC-IV		
Variables	**Coagulopathy ($n = 493$)**	**Non-Coagulopathy ($n = 506$)**	***P* Value**
Demographics			
Age (y), median [Q1, Q3]	67.00 (52.00, 80.00)	66.00 (48.00, 82.00)	0.809
Male, *n* (%)	317 (62.65)	343 (69.57)	0.025
Race, *n* (%)			
Black	24 (4.87)	28 (5.53)	
White	294 (59.63)	304 (60.08)	
Hispanic	14 (2.84)	16 (3.16)	
Asian	15 (3.04)	10 (1.98)	
Others	146 (29.61)	148 (29.25)	
BMI (kg/m^2), median [Q1, Q3]	26.25 (23.03, 29.80)	26.12 (23.10, 30.10)	0.914
Family history of stroke, *n* (%)	19 (3.85)	5 (0.99)	0.006
Coexisting disorders, *n* (%)			
Myocardial infarction	69 (14.00)	23 (4.55)	<0.001
Congestive heart failure	122 (24.75)	44 (8.70)	<0.001
Peripheral vascular disease	39 (7.91)	20 (3.95)	0.012
Cerebrovascular disease	98 (19.88)	75 (14.82)	0.043
Dementia	22 (4.46)	45 (8.89)	0.008
Chronic pulmonary disease	75 (15.21)	61 (12.06)	0.173
Rheumatic disease	11 (2.23)	5 (0.99)	0.189
Peptic ulcer disease	12 (2.43)	5 (0.99)	0.128
Diabetes	108 (21.91)	122 (24.11)	0.452
Paraplegia	47 (9.53)	62 (12.25)	0.202
Renal disease	73 (14.81)	42 (8.30)	0.002
Malignant cancer	36 (7.30)	10 (1.98)	<0.001
Severe liver disease	23 (4.67)	0 (0.00)	<0.001
Metastatic solid tumor	10 (2.03)	2 (0.40)	0.038
AIDS	2 (0.41)	2 (0.40)	1.000
CCI, median [Q1, Q3]	5.00 (3.00, 7.00)	4.00 (2.00, 6.00)	<0.001
Vital signs (1st 24h)			
Temperature (°C), median [Q1, Q3]	37.10 (36.70, 37.50)	37.20 (36.90, 37.52)	0.017
MAP (mmHg), median [Q1, Q3]	79.00 (73.00, 86.00)	81.00 (76.00, 88.00)	<0.001
Heart rate (min), median [Q1, Q3]	86.00 (76.00, 99.00)	84.00 (73.00, 95.00)	0.004
Respiratory rate (min), median [Q1, Q3]	19.00 (17.00, 22.00)	18.00 (16.00, 20.00)	<0.001
Laboratory findings (1st 24h)			
RBC (10^9/L), median [Q1, Q3]	3.40 (3.00, 3.80)	3.80 (3.30, 4.20)	<0.001
WBC ($\times 10^9$/L), median [Q1, Q3]	11.60 (8.33, 14.80)	12.30 (9.50, 15.00)	0.029
HGB (g/dL), median [Q1, Q3]	11.00 (9.00, 12.00)	12.00 (10.00, 13.00)	<0.001
PLT ($\times 10^9$/L), median [Q1, Q3]	166.50 (119.00, 224.75)	219.00 (178.00, 265.00)	<0.001
RDW (%), median [Q1, Q3]	17.20 (14.50, 47.80)	15.50 (13.60, 44.98)	<0.001
HCT (%), median [Q1, Q3]	31.90 (27.83, 35.77)	35.10 (31.20, 38.40)	<0.001
APTT (s), median [Q1, Q3]	31.40 (27.70, 38.20)	27.60 (25.70, 30.17)	<0.001
PT (s), median [Q1, Q3]	15.40 (13.40, 17.90)	12.90 (12.00, 13.80)	<0.001
INR, median [Q1, Q3]	1.40 (1.20, 1.60)	1.20 (1.10, 1.20)	<0.001
pH, median [Q1, Q3]	7.39 (7.34, 7.43)	7.40 (7.37, 7.44)	0.001
Bicarbonate (mmol/L), median [Q1, Q3]	22.50 (20.00, 25.00)	23.30 (21.00, 25.00)	0.002
Lactate (mmol/L), median [Q1, Q3]	1.80 (1.20, 2.60)	1.50 (1.00, 2.12)	< 0.001
BE (mEq/L), median [Q1, Q3]	−0.71 (-3.00, 1.00)	0.00 (-1.50, 1.50)	< 0.001
Anion gap (mmol/L), median [Q1, Q3]	14.80 (12.80, 16.70)	14.50 (13.00, 16.30)	0.467
PaO_2 (mmHg), median [Q1, Q3]	141.48 (104.91, 191.65)	148.33 (103.25, 193.69)	0.560
$PaCO_2$ (mmHg), median [Q1, Q3]	38.33 (35.00, 42.67)	38.46 (35.00, 43.00)	0.784

(Continued)

TABLE 1 | Continued

Variables	MIMIC-IV		
	Coagulopathy ($n = 493$)	Non-Coagulopathy ($n = 506$)	*P* Value
FiO_2 (%), median [Q1, Q3]	50.00 (42.50, 60.00)	50.00 (40.00, 57.50)	0.025
PaO_2/FiO_2, median [Q1, Q3]	286.29 (208.26, 372.00)	313.72 (227.25, 413.23)	0.008
Chloride (mmol/L), median [Q1, Q3]	105.50 (102.00, 109.30)	104.50 (101.00, 108.00)	0.001
Calcium (mmol/L), median [Q1, Q3]	8.30 (7.80, 8.70)	8.50 (8.00, 8.90)	<0.001
Sodium, (mmol/L), median [Q1, Q3]	140.00 (137.00, 142.80)	140.00 (137.00, 141.80)	0.049
Potassium (mmol/L), median [Q1, Q3]	4.10 (3.80, 4.40)	4.00 (3.80, 4.30)	0.197
Glucose (mmol/L), median [Q1, Q3]	141.00 (116.00, 166.00)	133.00 (114.50, 159.00)	0.035
CRE (mg/dL), median [Q1, Q3]	1.00 (0.70, 1.30)	0.90 (0.70, 1.10)	<0.001
BUN (mg/dL), median [Q1, Q3]	17.50 (12.30, 26.70)	15.00 (11.00, 20.00)	<0.001
Urine output (mL), median [Q1, Q3]	1668.00 (1078.00, 2462.50)	1875.00 (1250.00, 2673.75)	0.018
Type of injury, *n* (%)			
Subarachnoid hemorrhage	175 (35.50)	162 (32.02)	0.273
Cranial extradural hematoma	18 (3.65)	16 (3.16)	0.801
Cerebral contusion	74 (15.01)	124 (24.51)	< 0.001
Therapy strategy (1st 24h), *n* (%)			
MV	436 (88.44)	401 (79.25)	< 0.001
Blood Transfusion	29 (5.88)	5 (0.99)	< 0.001
Hyperosmolar therapy	46 (9.33)	63 (12.45)	0.139
Neurosurgical intervention	146 (29.61)	153 (30.24)	0.884
Scoring system			
GCS	8.00 (5.00, 10.00)	8.00 (7.00, 10.00)	0.001
SOFA	7.00 (5.00, 10.00)	5.00 (4.00, 6.00)	< 0.001
APSIII	39.00 (31.00, 48.00)	35.00 (27.00, 43.00)	< 0.001

MIMIC-IV, Medical Information Mart for Intensive Care-IV; BMI, body mass index; AIDS, acquired immunodeficiency syndrome; CCI, Charlson comorbidity index; MAP, mean artery pressure; RBC, red blood cell; WBC, white blood cell; HGB, hemoglobin; PLT, platelet; RDW, red blood cell volume distribution width; RDW, red blood cell volume distribution width; HCT, hematocrit; APTT, activated partial thromboplastin time; PT, prothrombin time; INR, international normalized ratio; BE, buffer excess; CRE, creatinine; BUN, blood urea nitrogen; MV, mechanical ventilation; GCS, Glasgow coma score; SOFA, sepsis related organ failure assessment; APSIII acute physiology score III; Blood Transfusion: defined as RBC, Plasma, PLT product administered; Hyperosmolar therapy: defined as HTS or mannitol; Neurosurgical intervention: defined as craniectomy or ventriculostomy.

established, and the testing set obtained AUCs of 0.910, 0.867, 0.920, 0.924, 0.915, 0.881, and 0.917, respectively. Comparatively, Ada had the highest predictive performance among these seven models (AUC 0.924, 95% Confidence Interval (CI) 0.902 to 0.943), while NB had the poorest generalization ability (AUC 0.867, 95% CI 0.839 to 0.891). The decision curve compared the net benefit of the best model and alternative approaches for clinical decision making. As shown in **Figure 4**, the net benefits of the Ada model and XGB model surpassed those of other ML models, including NB for all threshold values, showing that these two models were more superior in predicting the TBI-IC in this cohort.

In the **Figure 5**, the fifth predictor variables in the ML models are shown. Each variable included in the study had varying importance over the TBI-IC relying on the ML approach. Overall, the coagulation profile (PLT, INR, PT) was the variable with relatively higher importance across all ML algorithms, followed by APTT, SOFA, and so forth.

Model Interpretation

We next used the Shiny to illustrate the impacts of key features on the coagulopathy prediction model in individual patients. As shown in **Figure 6**, the information of one patient was input into the model: PLT (186×10^9/L), INR (1.1), PT (12s), APTT (29s), SOFA (4), RDW (44%), no congestive heart failure, RBC (3.9×10^9/L), CRE (8.7 mg/dL), BUN (24 mg/dL), sodium (142.3 mmol/L), HCT (39.2%), no myocardial infarction, no blood transfusion, HGB (14 g/dl). The model analyzed that the risk of coagulopathy in this patient was 82.10%, indicating that the probability of coagulopathy for the patients was high, and precaution measures were recommended.

DISCUSSION

Altered hemostasis and hemorrhagic progression are substantial challenges in the clinical management of TBI. Patients with TBI-IC were at a high risk of death over those with normal coagulation. Notably, studies elucidating the rapid prediction of TBI-IC, are warranted. In this sense, our study developed and validated ML models, providing an accurate predictive tool for coagulopathy in TBI patients. Specifically, seven ML models (NNET, NB, GBM, Ada, RF, BT and XGB) were used to predict TBI-IC using variables frequently used in clinical practice. Concerning the predictive performance, the Ada outperformed

TABLE 2 | Prediction performance of the machine learning models in the testing set.

Model	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC	95% CI
NNET	0.851	0.733	0.932	0.882	0.835	0.910	(0.886, 0.930)
NB	0.814	0.586	0.971	0.933	0.772	0.867	(0.839, 0.891)
GBM	0.848	0.800	0.881	0.823	0.864	0.920	(0.897, 0.939)
Ada	0.855	0.730	0.942	0.897	0.834	0.924	(0.902, 0.943)
RF	0.862	0.797	0.908	0.857	0.866	0.915	(0.892, 0.935)
BT	0.835	0.747	0.896	0.832	0.837	0.881	(0.854, 0.904)
XGB	0.859	0.744	0.939	0.895	0.841	0.917	(0.894, 0.936)

PPV, positive predictive values; NPV, negative predictive values, AUC, area under the curve; CI, confidence interval; NNET, artificial neural network; NB, naïve bayes; GBM, gradient boosting machine; Ada, adapting boosting; RF, random forest; BT, bagged trees; XGB, eXtreme Gradient Boosting.

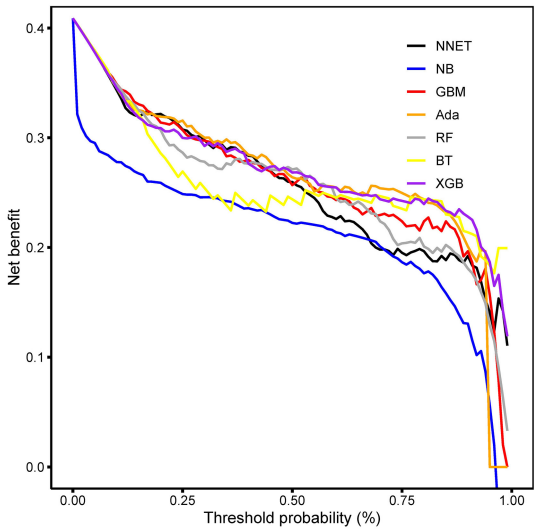


FIGURE 4 | Decision curve analysis. The net benefits of the Ada and XGB are relatively larger over a range of threshold probability values compared with those of other ML models. Ada, adapting boosting; XGB, eXtreme Gradient Boosting; ML, machine learning. NNET, arti?cial neural network; NB, naïve bayes; GBM, gradient boosting machine; RF, random forest; BT, bagged trees.

the remaining models. Moreover, results from the DCA indicated that the Ada and XGB models had higher net benefits over a range of threshold probability values than other models. It is remarkable that this study combined preoperative characteristic, comorbidities, and laboratory findings other than coagulopathy profile to establish a prediction model.

To help surgeons use the model, a calculator was developed, which provided a user-friendly interface. After entering the variables, the incidence of TBI-IC will be shown. The explanation of the ML model at the individual level was consistent with the aforementioned explanations at the feature level, and gratifyingly, the black-box concern was further mitigated to a certain extent. Notably, these results facilitated correct clinical decisions, and more importantly, timely treatment strategy.

A previous study conducted by Cosgriff et al. (20) developed a simple score to predict traumatic brain injury-induced coagulopathy (TIC) using four binary predictors [systolic blood pressure<70 mm Hg, temperature <34°C, pH <7.1, and Injury Severity Score (ISS) >25]. However, due to the fact that the ISS cannot be obtained at the time of decision making, the application of such a score was limited. To predict TIC more accurately, two scores have been developed by prehospital information (21, 22). Mitraet al.'s score used 5 predictors (entrapment; systolic blood pressure < 100 mm Hg; temperature < 34°C; suspected abdominal or pelvic injury; and chest decompression), whereas Peltan et al.'s score employed 6 predictors (age, injury mechanism, prehospital shock index> = 1, GCS, and need for prehospital tracheal intubation and/or Cardiopulmonary Resuscitation (CPR)) (21, 22). Nevertheless, in new patients, both scores achieved only moderate performance, with sensitivity <30%. Additionally, the Trauma Induced Coagulopathy Clinical Score (TICCS) employed three components, including general severity, blood pressure, and extent of significant injuries to predict TIC (23). A major limitation of above scores was that much of the prognostic potential of available information was lost through limiting the number of predictors and dichotomizing continuous variables. Consequently, a novel predictive model for early-identification of TIC was established (Predictors: heart rate, systolic blood pressure, temperature, hemothorax, Focused Assessment with Sonography for Trauma (FAST) result, unstable pelvic fracture, long bone fracture, GCS, lactate, base deficit, pH, mechanism of injury, energy) (24). However, one point worth noting was that previous study focused on the entire trauma patient, not TBI patients in particular, which added confusion to some extent.

By interpreting the full model, it was found that many clinical variables can contribute to predict the risk of TBI-IC. In this study, coagulopathy profile (INR, PT, APTT) was found to be the most important variable in predicting TBI-IC, followed by SOFA, blood routine test (PLT, RBC, HCT, HGB, RDW), renal function (BUN and CRE), comorbidities (congestive heart failure, myocardial infarction) and so forth. Among the fifteen included variables, the SOFA was an important predictor. SOFA is an indicator to describe multiple organ dysfunction, including respiratory system, nervous system, cardiovascular system, liver, coagulation and kidney (25). Potential mechanisms may include the fact that SOFA scores are more likely to indicate liver failure or cardiovascular failure. Those organ

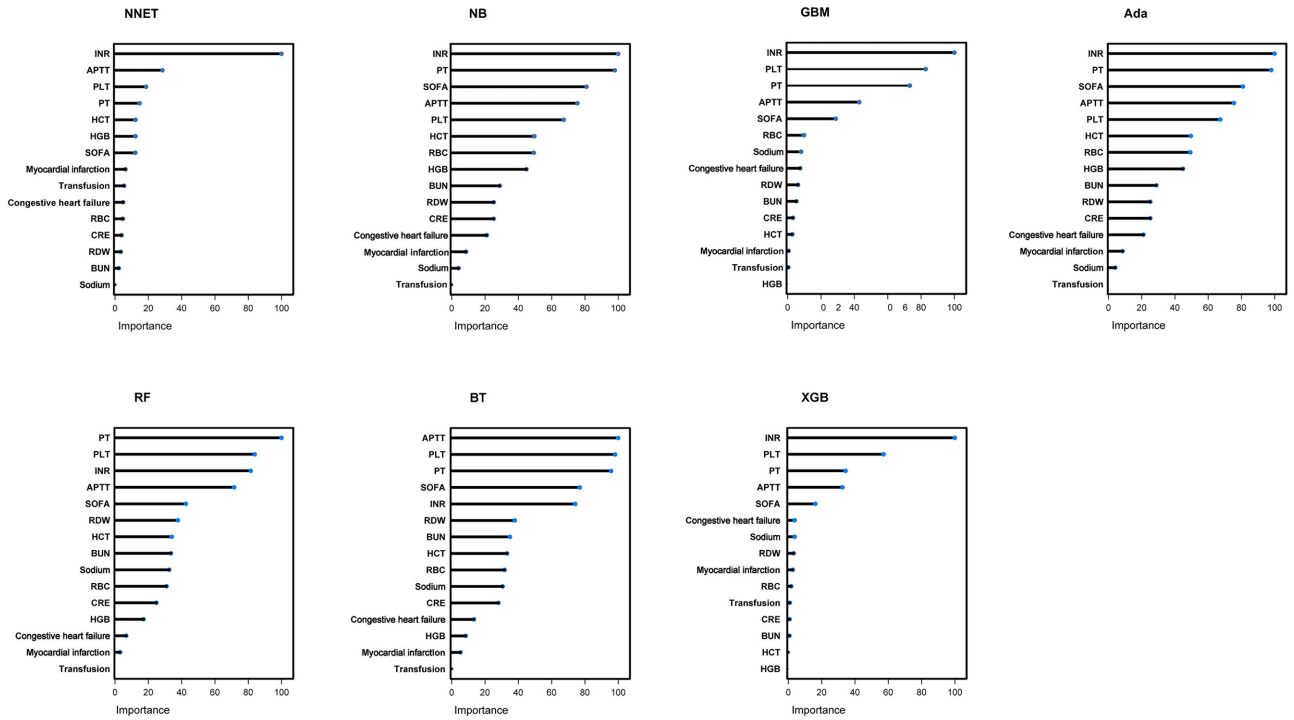


FIGURE 5 | Variable importance in seven different ML models. ML, machine learning; NNET, artificial neural network; NB, naïve bayes; GBM, gradient boosting machine; Ada, adapting boosting; RF, random forest; BT, bagged trees; XGB, eXtreme Gradient Boosting; INR, international normalized ratio; PT, prothrombin time; SOFA, sepsis related organ failure assessment; APTT, activated partial thromboplastin time; PLT, platelet; HCT, hematocrit; RBC, red blood cell; HGB, hemoglobin; BUN, blood urea nitrogen; RDW, red blood cell volume distribution width; CRE, creatinine.

failures have a high tendency to bleed, and subsequently leading to coagulopathy (26).

In this study, PLT, RBC, HCT, HGB and RDW were important predictors of TBI-IC. In a prospective observational study conducted by Davis PK et al. (27), PLT dysfunction was an early marker for TBI-IC. Potential mechanism included the blood dilution arised from the use of coagulation factor products (28). Nevertheless, we cannot exclude the likelihood that the blood coagulation system was activated by the continuous bleeding itself (29).

RDW, a parameter of red blood cell volume, measures the variability in size of circulating erythrocytes. Although primarily used to diagnose different types of anemias, the RDW was also associated with various thrombotic disease processes including venous thromboembolism (VTE) (30, 31).

Although the underlying mechanism is unclear, it is speculated that inflammatory factors destroy the vascular endothelial integrity, subsequently changing the glycoprotein and ion channel structure of the erythrocyte membrane (32, 33). Consequently, the deformability of the RBC is reduced, in turn, further enables endothelial damage to increase, causing the release of tissue factors that activate the coagulation pathway and triggers disseminated intravascular coagulation (DIC) (34).

In this study we found that renal function indicators (BUN and CRE) can help to indicate the risk of TBI-IC. Similarly, a ML model developed by Zhao QY et al. also identified renal function, including urine output and CRE to predict sepsis-induced coagulopathy (SIC) (35). It is worth noting that renal dysfunction has been associated with both thrombotic and hemorrhagic complications (36, 37). Potential mechanism included less adenosine diphosphate (ADP) and serotonin storage in PLT of patients with renal dysfunction (38, 39). Taken together, the force of impact at the time of TBI can cause shearing of large and small vessels, and result in subdural, subarachnoid, or intracerebral hemorrhages, or a combination of different types. TBI-associated factors might then alter the intricate balance between bleeding and thrombosis formation, leading to coagulopathy (9). Indeed, the complex interactions between the PLT dysfunction, changes in endogenous procoagulant, anticoagulant factors, endothelial cell activation, hypoperfusion, and inflammation related to TBI-IC remain to be elucidated (9, 40, 41).

The strengths of this study lied in the fact that it applied modern ML approaches to predict TBI-IC. It is worth noting that early and accurate prediction of TBI-IC can provide more time for clinicians to adjust corresponding treatment strategies. For example, this model is applicable if detailed medical history is not available for intubated severe head-injured ICU patient. Furthermore, given the heterogeneity of TBI-IC phenotypes (bleeding/thrombotic tendencies), timely treatment strategy would still require investigation and further testing to determine the type and therefore appropriate treatment. Furthermore, it was based on a real-world data with multicenter

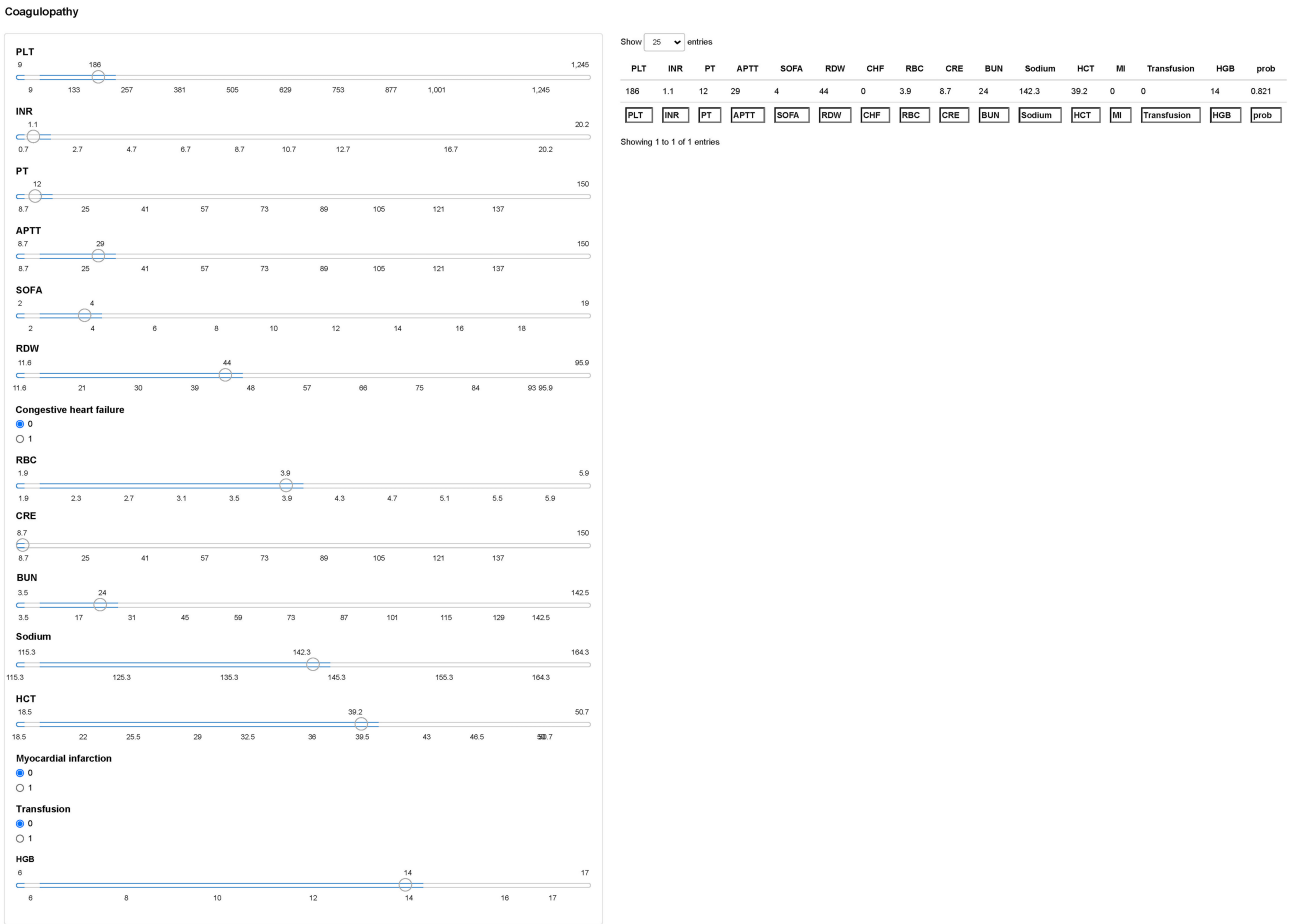


FIGURE 6 | Examples of website usage. Entering the input value determined the coagulopathy and displayed how each value contributed to the prediction. PLT, platelet; INR, international normalized ratio; PT, prothrombin time; APTT, activated partial thromboplastin time; SOFA, sepsis related organ failure assessment; RDW, red blood cell volume distribution width; RBC, red blood cell; CRE, creatinine; BUN, blood urea nitrogen; HCT, hematocrit; HGB, hemoglobin.

and external validation, which heighted the reliability of the performance of ML models. Besides, all the information in this dataset was coded independently of the practitioner, making it a reliable source.

Our study had limitations, consistent with those inherent to many large administrative database studies. First, only TBI-IC adults in ICUs were included, while TBI-IC children and hospitalized TBI-IC cases were not analyzed. Nevertheless, in light of the immaturity of the coagulation system in children, more research is indeed required. Second, derived from the ICU participants, the results of our study cannot be generalized to other population, and we did not obtain information including laboratory testing and interventions before ICU admission, which may cause confounders to some extent. Although our models can screen out patients who are at a high risk of TBI-IC, it is the surgeons who decide the administration of anticoagulant therapy. Usually, the interventions are time sensitive and need to occur early after admission, starting in the emergency department. Third, some new coagulation markers, for example, thrombin-antithrombin-III complex and plasmin-α2-antiplasmin complex, are useful in coagulopathy diagnosis (42, 43). Nevertheless, these indicators were not recorded in the MIMIC-IV and eICU database. This was also the case for viscoelastic coagulation testing [Thrombelastograghy (TEG), Rotational thromboelastometry (ROTEM), ClotPro]. Although these testings can provide detailed coagulopathy diagnosis rapidly and have multiple advantages over the traditional plasma-based coagulation tests (PT, APTT, INR), unfortunately, the above indicators were not included in these two databases. Fourth, we did not obtain the results of cranial Computer Tomography (CT) scans in this study, consequently, the original Corticosteroid Randomization After Significant Head Injury (CRASH)-CT score was not available. Moreover, as an administrative database, there was possibility for misclassification of TBI, to reduce bias caused by imprecise coding, we adopted the extensively used ICD-9, 10 codes. Fifth, as with all potential retrospective studies, there was a potential for unmeasured confounders, causing selection bias. Another major limitation worth noting was the changing nature of the variables in a critically ill patient from time of injury and right

throughout the continuum of care to ICU discharge. The nature of the retrospective database did not allow for correction for when measurements were taken in relation to the time of injury. Lastly, although our study deeply explored the coagulopathy of TBI in the ICU settings, other outcomes, such as long-term incidence, are also needed further investigation.

CONCLUSIONS

In general, the present study suggested that some important features were potentially related to the TBI-IC. The ML model processed large number of variables and subsequently discriminated TBI patients who would and would not develop coagulopathy, facilitating the implement of timely yet efficient treatments. In the future, further validation regarding its clinical application value will become a necessity.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: These data can be found here: https://mimic-iv.mit.edu/; https://eicu-crd.mit.edu/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

FY, CP, and LP: Conception and design. YL: Administrative support. JW: Collection and assembly of data. FY and CP: Data analysis and interpretation. All authors: Manuscript writing and final approval.

ACKNOWLEDGMENTS

We would like to thank the Massachusetts Institute of Technology and the Beth Israel Deaconess Medical Center for the MIMIC project. We also would like to thank the Philips eICU Research Institute and Philips Healthcare for their contribution to the eICU-CRD project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.792689/full#supplementary-material

Supplementary Table 1 | Missing number (%) for included variables in the dataset.

Supplementary Table 2 | Baseline characteristics of the eICU-CRD cohorts.

REFERENCES

1. Iaccarino C, Carretta A, Nicolosi F, Morselli C. Epidemiology of severe traumatic brain injury. *J Neurosurg Sci.* (2018) 62:535–41. doi: 10.23736/S0390-5616.18.04532-0
2. Hoyt DB, A. clinical review of bleeding dilemmas in trauma. *Semin Hematol.* (2004) 41:40–3. doi: 10.1053/j.seminhematol.2003.11.009
3. Harhangi BS, Kompanje EJ, Leebeek FW, Maas AI. Coagulation disorders after traumatic brain injury. *Acta Neurochirurgica.* (2008) 150:165–75. doi: 10.1007/s00701-007-1475-8
4. Maegele M, Schöchl H, Menovsky T, Maréchal H, Marklund N, Buki A, et al. Coagulopathy and haemorrhagic progression in traumatic brain injury: advances in mechanisms, diagnosis, and management. *Lancet Neurol.* (2017) 16:630–47. doi: 10.1016/S1474-4422(17)30197-7
5. Delano MJ, Rizoli SB, Rhind SG, Cuschieri J, Junger W, Baker AJ, et al. Prehospital resuscitation of traumatic hemorrhagic shock with hypertonic solutions worsens hypocoagulation and hyperfibrinolysis. *Shock (Augusta, Ga).* (2015) 44:25–31. doi: 10.1097/SHK.0000000000000368
6. Greuters S, van den Berg A, Franschman G, Viersen VA, Beishuizen A, Peerdeman SM, et al. Acute and delayed mild coagulopathy are related to outcome in patients with isolated traumatic brain injury. *Critical Care.* (2011) 15:R2. doi: 10.1186/cc9399
7. Samuels JM, Moore EE, Silliman CC, Banerjee A, Cohen MJ, Ghasabyan A, et al. Severe traumatic brain injury is associated with a unique coagulopathy phenotype. *J Trauma Acute Care Surg.* (2019) 86:686–93. doi: 10.1097/TA.0000000000002173
8. Talving P, Benfield R, Hadjizacharia P, Inaba K, Chan LS, Demetriades D. Coagulopathy in severe traumatic brain injury: a prospective study. *J Trauma.* (2009) 66:55–61. doi: 10.1097/TA.0b013e318190c3c0
9. Laroche M, Kutcher ME, Huang MC, Cohen MJ, Manley GT. Coagulopathy after traumatic brain injury. *Neurosurgery.* (2012) 70:1334–45. doi: 10.1227/NEU.0b013e31824d179b
10. Chen H, Xue LX, Guo Y, Chen SW, Wang G, Cao HL, et al. The influence of hemocoagulation disorders on the development of posttraumatic cerebral infarction and outcome in patients with moderate or severe head trauma. *Biomed Res Int.* (2013) 2013:685174. doi: 10.1155/2013/685174
11. Sun Y, Wang J, Wu X, Xi C, Gai Y, Liu H, et al. Validating the incidence of coagulopathy and disseminated intravascular coagulation in patients with traumatic brain injury–analysis of 242 cases. *Br J Neurosurg.* (2011) 25:363–8. doi: 10.3109/02688697.2011.552650
12. Beam AL, Kohane IS. Big data and machine learning in health care. *Jama.* (2018) 319:1317–8. doi: 10.1001/jama.2017.18391
13. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Critical Care.* (2019) 23:112. doi: 10.1186/s13054-019-2411-z
14. Zhang Z. Predictive analytics in the era of big data: opportunities and challenges. *Ann Translat Med.* (2020) 8:68. doi: 10.21037/atm.2019.10.97
15. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation.* (2000) 101:E215–20. doi: 10.1161/01.CIR.101.23.e215
16. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific Data.* (2018) 5:180178. doi: 10.1038/sdata.2018.178
17. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med.* (2015) 12:e1001885. doi: 10.1371/journal.pmed.1001885

18. Chang T, Yan X, Zhao C, Zhang Y, Wang B, Gao L. Risk factors and neurologic outcomes in patients with traumatic brain injury and coagulopathy within 72 h after surgery. *Neuropsychiatr Dis Treat.* (2021) 17:2905–13. doi: 10.2147/NDT.S323897
19. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* (2006) 26:565–74. doi: 10.1177/0272989X06295361
20. Cosgriff N, Moore EE, Sauaia A, Kenny-Moynihan M, Burch JM, Galloway B. Predicting life-threatening coagulopathy in the massively transfused trauma patient: hypothermia and acidoses revisited. *J Trauma.* (1997) 42:857–61. doi: 10.1097/00005373-199705000-00016
21. Mitra B, Cameron PA, Mori A, Maini A, Fitzgerald M, Paul E, et al. Early prediction of acute traumatic coagulopathy. *Resuscitation.* (2011) 82:1208–13. doi: 10.1016/j.resuscitation.2011.04.007
22. Peltan ID, Rowhani-Rahbar A, Vande Vusse LK, Caldwell E, Rea TD, Maier RV, et al. Development and validation of a prehospital prediction model for acute traumatic coagulopathy. *Critical Care.* (2016) 20:371. doi: 10.1186/s13054-016-1541-9
23. Tonglet ML, Minon JM, Seidel L, Poplavsky JL, Vergnion M. Prehospital identification of trauma patients with early acute coagulopathy and massive bleeding: results of a prospective non-interventional clinical trial evaluating the Trauma Induced Coagulopathy Clinical Score (TICCS). *Critical Care.* (2014) 18:648. doi: 10.1186/s13054-014-0648-0
24. Perkins ZB, Yet B, Marsden M, Glasgow S, Marsh W, Davenport R, et al. Early identification of trauma-induced coagulopathy: development and validation of a multivariable risk prediction model. *Ann Surg.* (2020). doi: 10.1097/SLA.0000000000003771 [Epubh ahead of print].
25. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Inten Care Med.* (1996) 22:707–10. doi: 10.1007/BF01709751
26. Akaraborworn O, Chaiwat O, Chatmongkolchart S, Kitsiripant C, Chittawatanarat K, Morakul S, et al. Prediction of massive transfusion in trauma patients in the surgical intensive care units (THAI-SICU study). *Chinese journal of traumatology = Zhonghua chuang shang za zhi.* (2019) 22:219–22. doi: 10.1016/j.cjtee.2019.04.004
27. Davis PK, Musunuru H, Walsh M, Cassady R, Yount R, Losiniecki A, et al. Platelet dysfunction is an early marker for traumatic brain injury-induced coagulopathy. *Neurocrit Care.* (2013) 18:201–8. doi: 10.1007/s12028-012-9745-6
28. Maegele M, Lefering R, Yucel N, Tjardes T, Rixen D, Paffrath T, et al. Early coagulopathy in multiple injury: an analysis from the German Trauma Registry on 8724 patients. *Injury.* (2007) 38:298–304. doi: 10.1016/j.injury.2006.10.003
29. Brohi K, Cohen MJ, Davenport RA. Acute coagulopathy of trauma: mechanism, identification and effect. *Curr Opin Crit Care.* (2007) 13:680–5. doi: 10.1097/MCC.0b013e3282f1e78f
30. Ellingsen TS, Lappegård J, Skjelbakken T, Brækkan SK, Hansen JB. Red cell distribution width is associated with incident venous thromboembolism (VTE) and case-fatality after VTE in a general population. *Thromb Haemost.* (2015) 113:193–200. doi: 10.1160/TH14-04-0335
31. Zöller B, Melander O, Svensson P, Engström G. Red cell distribution width and risk for venous thromboembolism: a population-based cohort study. *Thromb Res.* (2014) 133:334–9. doi: 10.1016/j.thromres.2013.12.013
32. Piagnerelli M, Boudjeltia KZ, Brohee D, Piro P, Carlier E, Vincent JL, et al. Alterations of red blood cell shape and sialic acid membrane content in septic patients. *Crit Care Med.* (2003) 31:2156–62. doi: 10.1097/01.CCM.0000079608.00875.14
33. Lasocki S, Lefebvre T, Mayeur C, Puy H, Mebazaa A, Gayat E. Iron deficiency diagnosed using hepcidin on critical care discharge is an independent risk factor for death and poor quality of life at one year: an observational prospective study on 1161 patients. *Critical Care.* (2018) 22:314. doi: 10.1186/s13054-018-2253-0
34. Levi M, van der Poll T. Inflammation and coagulation. *Critical Care Med.* (2010) 38:S26–34. doi: 10.1097/CCM.0b013e3181c98d21
35. Zhao QY, Liu LP, Luo JC, Luo YW, Wang H, Zhang YJ, et al. A machine-learning approach for dynamic prediction of sepsis-induced coagulopathy in critically ill patients with sepsis. *Front Med.* (2020) 7:637434. doi: 10.3389/fmed.2020.637434
36. Boccardo P, Remuzzi G, Galbusera M. Platelet dysfunction in renal failure. *Semin Thromb Hemost.* (2004) 30:579–89. doi: 10.1055/s-2004-835678
37. Iseki K, Kinjo K, Kimura Y, Osawa A, Fukiyama K. Evidence for high risk of cerebral hemorrhage in chronic dialysis patients. *Kidney Int.* (1993) 44:1086–90. doi: 10.1038/ki.1993.352
38. Sohal AS, Gangji AS, Crowther MA, Treleaven D. Uremic bleeding: pathophysiology and clinical risk factors. *Thromb Res.* (2006) 118:417–22. doi: 10.1016/j.thromres.2005.03.032
39. Gawaz MP, Dobos G, Späth M, Schollmeyer P, Gurland HJ, Mujais SK. Impaired function of platelet membrane glycoprotein IIb-IIIa in end-stage renal disease. *J Am Soc Nephrol JASN.* (1994) 5:36–46. doi: 10.1681/ASN.V5136
40. Chang R, Cardenas JC, Wade CE, Holcomb JB. Advances in the understanding of trauma-induced coagulopathy. *Blood.* (2016) 128:1043–9. doi: 10.1182/blood-2016-01-636423
41. Di Battista AP, Rizoli SB, Lejnieks B, Min A, Shiu MY, Peng HT, et al. Sympathoadrenal activation is associated with acute traumatic coagulopathy and endotheliopathy in isolated brain injury. *Shock.* (2016) 46:96–103. doi: 10.1097/SHK.0000000000000642
42. Gall LS, Davenport RA. Fibrinolysis and antifibrinolytic treatment in the trauma patient. *Curr Opin Anaesthesiol.* (2018) 31:227–33. doi: 10.1097/ACO.0000000000000561
43. Müller MC, Meijers JC, Vroom MB, Juffermans NP. Utility of thromboelastography and/or thromboelastometry in adults with sepsis: a systematic review. *Critical Care.* (2014) 18:R30. doi: 10.1186/cc13721

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 24 December 2021
doi: 10.3389/fmed.2021.793230

Machine Learning Model for Predicting Acute Respiratory Failure in Individuals With Moderate-to-Severe Traumatic Brain Injury

Rui Na Ma[1†], Yi Xuan He[2†], Fu Ping Bai[3], Zhi Peng Song[2], Ming Sheng Chen[2] and Min Li[2]*

[1] Department of Pulmonary and Critical Care Medicine, The Second Affiliated Hospitals of Fourth Military Medical University, Xi'an, China, [2] Neurocritical Care Unit, Department of Neurosurgery, The Second Affiliated Hospitals of Fourth Military Medical University, Xi'an, China, [3] Department of Neurosurgery, Lin Fen Center Hospital, Lin Fen, China

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Arulselvi Subramanian,
All India Institute of Medical Sciences, India
Mohammad Ali Heidari Gorji,
Iran University of Medical Sciences, Iran

****Correspondence:***
Min Li
neursylm@fmmu.edu.cn

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *11 October 2021*
Accepted: *01 December 2021*
Published: *24 December 2021*

Citation:
Ma RN, He YX, Bai FP, Song ZP, Chen MS and Li M (2021) Machine Learning Model for Predicting Acute Respiratory Failure in Individuals With Moderate-to-Severe Traumatic Brain Injury. Front. Med. 8:793230. doi: 10.3389/fmed.2021.793230

Background: There is a high incidence of acute respiratory failure (ARF) in moderate or severe traumatic brain injury (M-STBI), worsening outcomes. This study aimed to design a predictive model for ARF.

Methods: Adult patients with M-STBI [3 $\leq$ Glasgow Coma Scale (GCS) $\leq$ 12] with a definite history of brain trauma and abnormal head on CT images, obtained from September 2015 to May 2017, were included. Patients with age >80 years or <18 years, multiple injuries with TBI upon admission, or pregnancy (in women) were excluded. Two models based on machine learning extreme gradient boosting (XGBoost) or logistic regression, respectively, were developed for predicting ARF within 48 h upon admission. These models were evaluated by out-of-sample validation. The samples were assigned to the training and test sets at a ratio of 3:1.

Results: In total, 312 patients were analyzed including 132 (42.3%) patients who had ARF. The GCS and the Marshall CT score, procalcitonin (PCT), and C-reactive protein (CRP) on admission significantly predicted ARF. The novel machine learning XGBoost model was superior to logistic regression model in predicting ARF [area under the receiver operating characteristic (AUROC) = 0.903, 95% CI, 0.834–0.966 vs. AUROC = 0.798, 95% CI, 0.697–0.899; $p < 0.05$].

Conclusion: The XGBoost model could better predict ARF in comparison with logistic regression-based model. Therefore, machine learning methods could help to develop and validate novel predictive models.

Keywords: acute respiratory failure, machine learning, XGBoost model, logistic regression, traumatic brain injury

INTRODUCTION

Acute respiratory failure (ARF) is a common pathophysiological result of pulmonary complications [pneumonia, neurogenic pulmonary edema, and acute respiratory distress syndrome (ARDS)] in moderate or severe traumatic brain injury (M-STBI), not only worsening outcomes, but also extending intensive care unit (ICU) and hospital stays and increasing the cost of hospital care (1–7).

Consequently, accurately predicting ARF risk may help to identify cases requiring intensive airway management. This would help to allocate resources efficiently and improve morbidity reduction by appropriately monitoring patients at risk.

With the rapid development of software, there is increasing use of machine learning algorithms. Especially, machine learning methods have been applied in medicine with excellent results, deriving predictive algorithms for multiple conditions (8–15). While traditional predictive models employ selected parameters, machine learning methods easily include multiple clinical parameters (16).

Although some predictive score systems or risk calculators have been developed by previous studies for the prediction of pulmonary complications (3, 5, 9, 13, 17), to date, studies assessing RF feature selection and machine learning algorithms are rare in the M-STBI population.

We hypothesized that supervised machine learning could help to develop models for better predicting single ARF occurrence upon M-STBI compared with routine statistical models. Therefore, this study aimed to utilize a machine learning model for developing and validating an ARF predictive model, termed extreme gradient boosting (XGBoost), which was compared to a conventional logistic regression model for effectiveness.

MATERIALS AND METHODS

Data Source

Model development and internal validation were based on a large TBI database, which consists of data of patients admitted to the department of neurosurgery in the Second Affiliated Hospital of Fourth Military Medical University, China, from September 2015 to May 2017. This trial had approval from the Institutional Ethics Board of the Second Affiliated Hospital of Fourth Military Medical University (TDLL-KY-202110-09) and data reporting followed the guidelines included in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement (18).

Selection of Patient

Adult patients with M-STBI [3 $\leq$ Glasgow Coma Scale (GCS) $\leq$ 12] with a definite history of brain trauma and abnormal head on CT images, acquired from September 2015 to May 2017, were included in this study. Patients with age >80 years or <18 years, multiple injuries with TBI upon admission, or pregnancy (in women) were excluded from this study.

Data Collection and Outcome Definition

The medical records of the patients were carefully collected by three authors on separate occasions. Demographic parameters, clinical and laboratory variables, comorbidities, imaging features, and outcome variables were recorded. All the patients with M-STBI underwent the procedure of arterial blood gas (ABG) analysis within the day of admission; ABG was repeated, if oxygen saturation (SpO_2) <93% using a nasal catheter or mask oxygen inhalation for at least 5 min after suctioning oropharyngeal secretions. The primary endpoint of this study was ARF within 72 h of admission, which was defined as respiratory failure with partial pressure of oxygen (pO_2) <60 mm Hg and respiratory rate >30 breaths/min or respiratory distress for at least 5 min (19).

Predictors of ARF

Clinical and laboratory parameters recorded in the initial 48 h after ICU admission were examined for their capabilities of predicting ARF. For parameters measured many times, both the maxima and minima were examined. Age, gender, GCS, comorbidity, and imaging features including the Marshall CT score and severity scores of lung exudations (see **Tables S1, S2** for the details of the scores) were analyzed. In addition, laboratory data such as white blood cell (WBC) and neutrophil counts, neutrophil–lymphocyte ratio, C reactive protein (CRP), and procalcitonin (PCT) were included. In term of therapy, long-term sedation (sedation duration > 48 h) was examined. For predictor selection, the Akaike Information Criterion (AIC) was used for minimizing the possible collinearity of parameters from a given patient as well as overfitting (20).

This was a hypothesis-generating retrospective trial, with no sample size estimation, but including the totality of eligible patients in the database for statistical power maximization.

Missing Data

We aimed to reflect daily clinical routine where often not all the data are obtainable. To make our algorithms and study realistic, we decided not to correct for missing data, e.g., by imputation techniques and to perform the analysis using the available data only. While using imputation techniques to estimate missing variables have many merits in conventional statistics, it is less preferred in machine learning because it does not reflect the observed reality—at best a close approximation—and adds artificially introduced noise to the data. Moreover, there could be significant reasons why some data are missing, which could be linked to the outcome variable of interest. In such cases (and in a number of other scenarios), imputation obscures important relationships in the observed data or introduces artificial relationships altogether, which decreases the value of complex pattern recognition used in machine learning. For variables with missing values, we coded the missing value as zero and created the corresponding missing dummy (12).

Statistical Analysis

Continuous and categorical data were presented as median [interquartile range (IQR)] and number (percentage). Demographic characteristics between participants with and without ARF were compared by the Mann–Whitney *U* test or the chi-squared test.

The primary model of this study was the XGBoost gradient boosted tree model. XGBoost represents a tree ensemble technique building in a progressive fashion on the loss from weak decision tree base learners. It can learn rapidly and effectively from substantial data amounts, with a flexibility allowing learning even from missing data (21). After tuning the XGBoost model, parameters of the XGBoost model were finally max_depth = 7, subsample = 0.94, colsample_bytree = 0.83, nrounds = 100, learning rate (eta value) = 0.3, and gamma = 5.

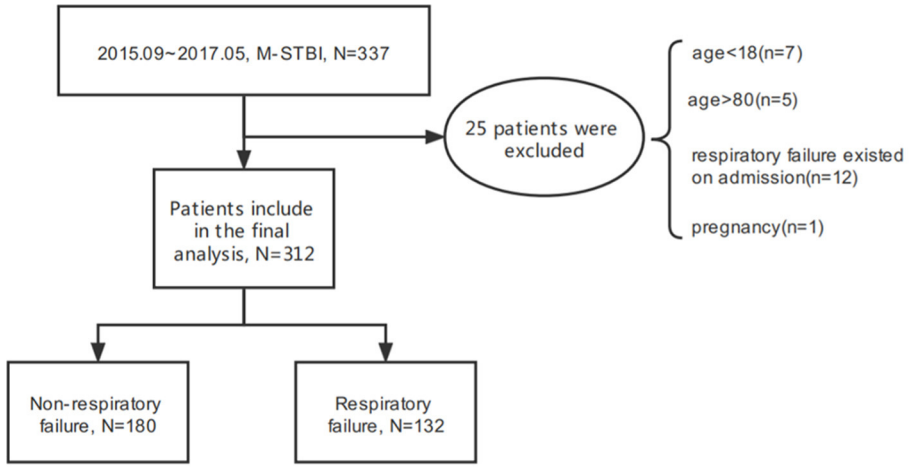


FIGURE 1 | Study flowchart.

For comparison, another model for predicting ARF occurrence was developed based on the multivariate logistic regression analysis.

As a comparison, a second model to predict the occurrence of ARF was created using the multivariate logistic regression model.

For comparison, model discrimination was assessed using the area under the receiver operating characteristic (AUROC) curve and the optimal cutoff value was calculated by Youden index. The confusion matrixes of the two models were created based on the optimal cutoff values to evaluates the accuracy, sensitivity, and specificity.

EmpowerStats (X&Y Solutions, Inc., Boston, MA, USA) and R version 3.4.2 (http://www.R-project.org) were utilized for data analysis. $p < 0.05$ was considered as statistically significant.

RESULTS

Patients

Between September 2015 and May 2017, 312 M-STBI cases hospitalized in the non-ICU (NICU) of the Second Affiliated Hospital of Fourth Military Medical University were examined. There were 232 males and 80 females. Of all patients, 132 (42.3%) patients had ARF (**Figure 1**). Characteristics of patients are given in detail in **Table 1**.

The ARF group included more individuals with smoking history (37.12 vs. 26.67%; $p = 0.049$) and chronic obstructive pulmonary disease (COPD) history (5.30 vs. 1.11%; $p = 0.029$) prior to ICU admission than the non-ARF group. Upon admission, the minimum GCS values (6.57 ± 2.68 vs. 8.63 ± 3.27 mmol/l; $p < 0.001$) were lower, while the Marshall CT scores (5.50 ± 0.95 vs. 4.70 ± 1.39; $p < 0.001$) and severity scores of bilateral lung exudations (83.33 vs. 66.67%; $p = 0.004$) were higher in ARF cases. ARF cases also showed elevated white blood cell count (14.87 ± 7.14 vs. 10.96 ± 5.16; $p < 0.001$), elevated neutrophil cell count (85.06 ± 9.47 vs. 78.27 ± 12.37; $p < 0.001$), lower neutrophil–lymphocyte ratio (5.66 ± 5.83 vs. 9.78 ± 10.11; $p < 0.001$), and higher CRP (57.10 ± 59.85 vs. 23.51 ± 31.19 mmol/l; $p < 0.001$) and PCT (2.54 ± 6.09 vs. 0.42 ± 1.13; $p = 0.002$) compared with the non-ARF group (**Table 1**).

The XGBoost Model

Extreme gradient boosting had an AUROC of 0.84 in the training set, with sensitivity and specificity of 0.71 and 0.84, respectively. Its precision was 0.78 (95% CI: 0.72–0.83). An error rate of 0.12 was obtained, indicating a correct prediction in roughly 78% of patients (**Table 2**).

In the test population, an AUROC of 0.90 was obtained for XGBoost, which had specificity and sensitivity of 0.85 and 0.78, respectively, indicating correct prediction of 29 of the 37 ARF cases in the test set. Meanwhile, 8 cases were incorrectly predicted [reflecting a precision rate of 0.82 (0.72, 0.90)]. The model had an error rate of 0.18, indicating correct outcome prediction in >81% of cases (**Table 2**).

Variables showing high predictive values were the GCS and the Marshall CT score, PCT, and CRP on admission. The GCS was the center factor of the XGBoost model because the gain of the GCS was the highest among all the variables (**Figure 2**). Other variables, e.g., long-term sedation and smoking history had low prediction power (**Figure 2**).

Logistic Regression Model

Baseline parameters for the ARF and non-ARF groups are shown in **Table 1**. Smoking and COPD history, the GCS and the Marshall CT score on admission, severity scores of lung exudations, long-term sedation, neutrophil cell count, WBC, neutrophil–lymphocyte ratio, PCT, and CRP showed associations with ARF occurrence in the univariate analysis ($p < 0.05$, **Table 1**). In the stepwise multivariate logistic regression analysis, bilateral lung exudations [odds ratio (OR), 3.435; 95% CI, 1.248–9.456], the Marshall CT score (OR for each 1 score increase, 1.078; 95% CI, 1.012–1.148), long-term sedation, increased WBC

TABLE 1 | Baseline characteristics of patients with moderate-to-severe traumatic brain injury with or without acute respiratory failure (ARF).

Exposure	Non-ARF (*N* = 180)	ARF (*N* = 132)	*P*-value
AGE	56.0 (45.8–66.0)	57.0 (49.0–66.0)	0.448
Sex (male/%)	129 (71.67%)	103 (78.03%)	0.203
Smoking	48 (26.67%)	49 (37.12%)	0.049
GCS	8.0 (6.0–11.0)	6.0 (5.0–8.0)	<0.001
Marshall score	5.0 (4.0–6.0)	6.0 (5.0–6.0)	<0.001
Scores of lung exudations			0.004
0	46 (25.56%)	16 (12.12%)	
1	14 (7.78%)	6 (4.55%)	
2	120 (66.67%)	110 (83.33%)	
Comorbidity			
Hypertension (*n*, %)	92 (51.11%)	69 (52.27%)	0.839
Diabetes	12 (6.67%)	7 (5.30%)	0.619
COPD	2 (1.11%)	7 (5.30%)	0.029
Cardiovascular disease	3 (1.67%)	7 (5.30%)	0.072
Long-term sedation	75 (41.67%)	102 (77.27%)	<0.001
White cell count, $\times 10^9$/L	10.3 (7.4–13.7)	13.6 (10.3–18.4)	<0.001
Neutrophil cell count, %	81.3 (73.2–87.0)	87.5 (82.5–90.7)	<0.001
Neutrophil-lymphocyte ratio	7.3 (3.0–11.6)	4.4 (1.4–8.0)	<0.001
CRP, mg/L	6.2 (5.0–28.0)	26.9 (5.0–97.7)	<0.001
Not recorded, n	104 (57.7%)	69 (52.3%)	
PCT, ng/mL	0.2 (0.2–0.3)	0.3 (0.2–1.7)	<0.001
Not recorded, n	96 (53.3%)	54 (40.9%)	
LOH	10.0 (7.0–16.0)	17.5 (11.0–28.0)	<0.001

Data are expressed as medians ± interquartile ranges and n (percentage), as appropriate.
GCS, Glasgow Coma Scale; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; PCT, procalcitonin.

TABLE 2 | The multivariate logistic regression model with stepwise variable selection.

Variables	OR (95% CL)	*P*-value
Smoking	2.092 (0.989–4.429)	0.054
Scores of lung exudations 1	0.988 (0.158–6.172)	0.990
Scores of lung exudations 2	3.435 (1.248–9.456)	0.017
WBC	1.078 (1.012–1.148)	0.020
GCS	0.788 (0.681–0.913)	0.002
Marshall score	1.706 (1.181–2.463)	0.004
Long-term sedation	6.293 (2.908–13.621)	0.001
PCT	1.121 (0.924–1.360)	0.249
CRP	1.014 (1.004–1.025)	0.007
[a]PCT (not recorded)	0.540 (0.245–1.190)	0.126

WBC, white blood cell; GCS, Glasgow Coma Scale; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; PCT, procalcitonin.
[a]PCT dummy variable for missing values.

(OR for each 1 × 10^9/L increase, 1.076; 95% CI, 1.181–2.463), and CRP (OR for each 1 mg/l increase, 1.014; 95% CI, 1.004–1.025) were associated with increased probability of ARF. On the contrary, the GCS (OR for each 1 score increase, 0.788; 95% CI, 0.681–0.913) was associated with decreased probability of ARF (**Table 2**).

The multivariate regression model was created based on the AIC-selected variables. It showed an AUROC of 0.943 in the training cohort, with a specificity of 0.946 and a sensitivity of 0.837 (**Table 3**). Its error rate was 11.6%. In the test population, AUROC was 0.792 and specificity and sensitivity were 0.913 and 0.667, respectively; its error rate approximated 15.6% (**Table 3**).

Model Performances

Area under the receiver operating characteristics were determined for assessing the discriminative abilities of both the models. XGBoost showed an elevated AUROC in comparison with the logistic regression model (AUROC, 0.902; 95% CI, 0.834–0.966 vs. 0.789; 95% CI, 0.688–0.891, $p < 0.05$; **Figure 3**). **Tables 3**, **4** describe the classification and confusion matrixes for both the models in predicting ARF.

DISCUSSION

Prediction and timely detection of ARF in patients with M-STBI are critical, crucially impacting M-STBI outcome (22, 23). This study developed a machine learning-based model to predict ARF occurrence in M-STBI, with multiple remarkable features. First, the model included readily available and reproducible parameters in the initial 48 h after admission. Second, after analyzing multiple interaction patterns among variables, the predominance of admission-related parameters (the GCS and the

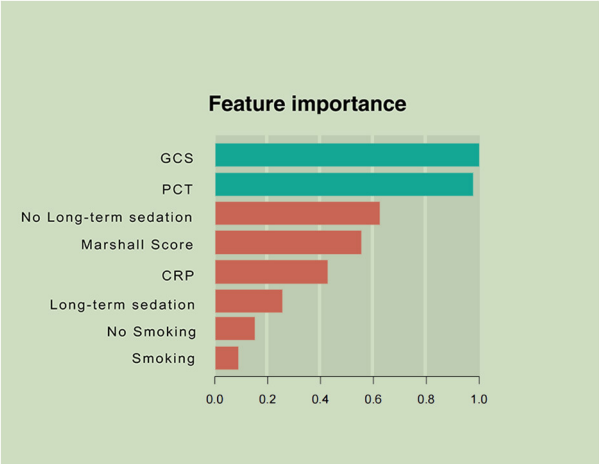


FIGURE 2 | Parameters by predictive value in the extreme gradient boosting (XGBoost) model. To predict acute respiratory failure (ARF) following moderate or severe traumatic brain injury, gradient boosting used various variables based on their importance in prediction modeling. In this analysis, the Glasgow Coma Scale (GCS) and inflammation-associated laboratory parameters upon admission had higher values in ARF prediction than other features of patient.

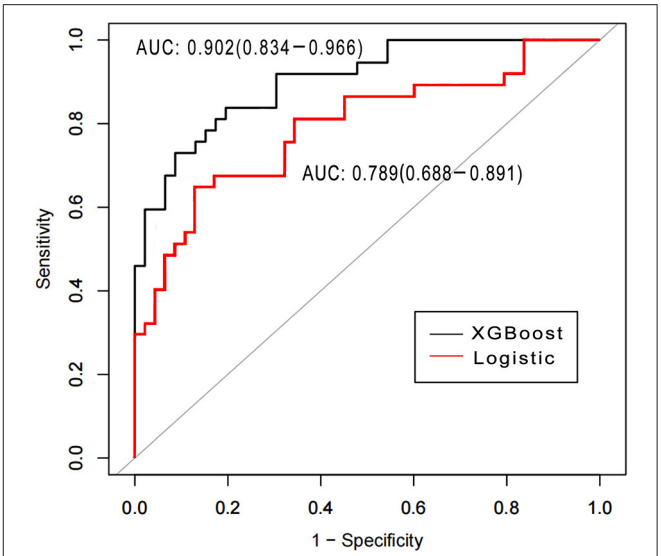


FIGURE 3 | Receiver operating characteristic curves for examining the discriminative powers of the XGBoost and the logistic regression models.

TABLE 3 | Confusion matrix for machine learning.

Training data				**Statistical analysis**	
Predicted true	0	1	Total	AU-ROC	0.840
0	112	28	140	Accuracy	0.782
1	22	67	89	Sensitivity	0.705
Total	134	95	229	Specificity	0.836
Test data				Statistical analysis	
Predicted true	0	1	Total	AU-ROC	0.902
0	39	8	47	Accuracy	0.820
1	7	29	36	Sensitivity	0.784
Total	46	37	83	Specificity	0.848

AU-ROC, area under the receiver operating characteristic.

TABLE 4 | Confusion matrix for conventional statistics.

Training data				**Statistical analysis**	
Predicted true	0	1	Total	AU-ROC	0.879
0	99	12	111	Accuracy	0.795
1	35	83	118	Sensitivity	0.874
Total	134	95	229	Specificity	0.739
Test data				Statistical analysis	
Predicted true	0	1	Total	AU-ROC	0.789
0	40	13	53	Accuracy	0.771
1	6	24	30	Sensitivity	0.649
Total	46	37	83	Specificity	0.870

AUROC, area under the receiver operating characteristic.

Marshall CT score, CRP, PCT, and long-term sedation; **Figure 2**) was most significant in determining the occurrence of ARF. Third, the novel model enhanced performance compared with the conventional logistic regression model.

This study first investigated ARF prediction in patients with M-STBI using machine learning methods. This new model had accuracy and AUROC of 0.83 and 0.90, respectively. Of greatest importance, sensitivity and specificity of 0.73 and 0.91, respectively, were obtained in the test cohort.

First, accurate detection of ARF in critically ill individuals with M-STBI is essential in performing intensive airway management and making decision with respect to invasive treatments such as tracheal intubation, invasive mechanical ventilation, and even tracheostomy. To date, reliable tools for timely predicting ARF in M-STBI are lacking. In this study, we demonstrated enlightened machine learning methods, including XGBoost, could provide a great deal of information obtainable from databases and promote the development and validation of better predictive models in comparison with conventional logistic regression techniques. The new model could help to stratify M-STBI cases right after ICU admission. Therefore, intensive airway management or invasive treatment could be more accurately provided to individuals with high odds of developing ARF to avoid long-term hypoxia, which is associated with increased morbidity and mortality in patients with M-STBI (24, 25). On the other hand, intensive airway management needs important human and material resources, while invasive treatment is related to complications and high medical costs. Thus, identifying individuals who could benefit from intensive airway management or invasive treatment are critical. However, this analysis provided

no high level of evidence with respect to the effectiveness of XGBoost. Further randomized controlled trials that compare therapies dependent on and independent of the predictive model should comprehensively examine its effectiveness.

Second, we aimed to design a model with easy implementation by neurosurgery residents and staff alike. Therefore, parameters easily available and reproducible upon admission were required and quantitative (blood test results, the GCS score, the Marshall CT score, etc.) and dichotomous (long-term sedation or not, smoking status, etc.) variables were selected.

Third, the XGBoost model showed that the GCS score, PCT, the Marshall CT score, CRP, and long-term sedation potentially predicted ARF in patients with M-STBI. Consistent with previous reports, the GCS score, PCT, and CRP were related to ARF in patients with M-STBI, suggesting the extent of TBI and severity of systematic inflammation (26–29). The GCS was center factor in the XGBoosting model shown in **Figure 2**, suggesting that the severity of brain injury was associated with ARF in patients with M-STBI significantly. The results agreed with clinical experience very well. However, to the best of our knowledge, the association between the Marshall CT score and ARF has not been confirmed. This study suggested that the Marshall CT score potentially predicted ARF. The explanation could be that the Marshall CT score can reflect the extent of brain injury based on neuroimaging, so the high Marshall CT score is associated with injury of brainstem centers of respiration or intracranial hypertension, which causes ARF easily. Moreover, both the logistic and XGBoost models showed that sedation (more than 48 h) was related to ARF. The results could be explained by the fact that sedation is an important tool for reducing intracranial pressure, which cannot be stopped until intracranial pressure returns to normal. Intracranial hypertension and respiratory depression caused by sedative drugs contribute to ARF (30, 31).

This study had many strengths. XGBoost modeling represents a new method not yet applied in respiratory failure studies of neurological critical patients. XGBoost modeling can learn swiftly with high efficiency from important data amounts and its high flexibility enables learning even from missing data (21). The XGBoost model had starkly higher predictive accuracy compared with the generalized linear model, being capable of capturing complex associations in data without requiring explicit high-order interactions and non-linear functions (12). Using such features, predictive models based on clinical and laboratory variables, which are easily available and reproducible upon admission, could be built. However, there were also limitations. First, as a hypothesis-generating study, external validation of the XGBoost model is important for confirming its usefulness. The XGBoost model developed in this study will be applied to the Medical Information Mart for Intensive Care (MIMIC)-IV for external validation in the next study. Second, because this was a retrospective study, missing data are inevitable in practice. For missing data, variables with >70% missing values were excluded from model construction. Thus, the sample sizes of the training ($n = 86$) and test ($n = 32$) sets were low especially in the logistic regression model. To some extent, missing data decreased the performance of the model. Third, this study only explored ARF within 48 h upon admission and a different time interval (e.g., >48 h following admission) was not studied.

CONCLUSION

In total, six major parameters related to ARF were screened to develop the XGBoost model with enhanced predictive value for ARF compared with the logistic regression model in patients with M-STBI.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the Second Affiliated Hospital, Fourth Military Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

ML and RM contributed to the study conception, design, and manuscript drafting. YH, MC, FB, and ZS contributed to the acquisition of data and analysis and interpretation of data. All authors approved the final version of the manuscript.

FUNDING

This study was funded by the Youth Innovation Fund of the Second Affiliated Hospital of Fourth Military Medical University (Grant No. 2021LCYJ026).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.793230/full#supplementary-material

REFERENCES

1. Khan NA, Quan H, Bugar JM, Lemaire JB, Brant R, Ghali AW. Association of postoperative complications with hospital costs and length of stay in a tertiary care center. *J Gen Intern Med*. (2006) 21:177–80. doi: 10.1007/s11606-006-0254-1
2. Gupta H, Gupta PK, Fang X, Miller WJ, Cemaj S, Forse RA, et al. Development and validation of a risk calculator predicting postoperative respiratory failure. *Chest*. (2011) 140:1207–15. doi: 10.1378/chest.11-0466
3. Canet J, Sabaté S, Mazo V, Gallart L, de Abreu MG, Belda J, et al. Development and validation of a score to predict postoperative respiratory failure in

a multicentre European cohort: a prospective, observational study. *Eur J Anaesthesiol*. (2015) 32:458–70. doi: 10.1097/EJA.0000000000000223
4. Abd El Aziz MA, Perry WR, Grass F, Mathis KL, Larson DW, Mandrekar J, et al. Predicting primary postoperative pulmonary complications in patients undergoing minimally invasive surgery for colorectal cancer. *Updates Surg*. (2020) 72:977–83. doi: 10.1007/s13304-020-00892-6
5. Ostermann RC, Joestl J, Tiefenboeck TM, Lang N, Platzer P, Hofbauer M. Risk factors predicting prognosis and outcome of elderly patients with isolated traumatic brain injury. *J Orthop Surg Res*. (2018) 13:277. doi: 10.1186/s13018-018-0975-y
6. Ho CH, Liang FW, Wang JJ, Chio CC, Kuo RJ. Impact of grouping complications on mortality in traumatic brain injury: a nationwide population-based study. *PLoS ONE*. (2018) 13:e0190683. doi: 10.1371/journal.pone.0190683
7. Krishnamoorthy V, Vavilala MS, Mills B, Rowhani-Rahbar A. Demographic and clinical risk factors associated with hospital mortality after isolated severe traumatic brain injury: a cohort study. *J Intensive Care*. (2015) 3:46. doi: 10.1186/s40560-015-0113-4
8. Cooper JP, Perkins JD, Warner PR, Shingina A, Biggins SW, Abkowitz JL, et al. Acute graft-versus-host disease following orthotopic liver transplantation: predicting this rare complication using machine learning. *Liver Transpl*. (2021). doi: 10.1002/lt.26318. [Epub ahead of print].
9. Mathioudakis NN, Abusamaan MS, Shakarchi AF, Sokolinsky S, Fayzullin S, McGready J, et al. Development and validation of a machine learning model to predict near-term risk of iatrogenic hypoglycemia in hospitalized patients. *JAMA Netw Open*. (2021) 4:e2030913. doi: 10.1001/jamanetworkopen.2020.30913
10. Mathis MR, Engoren MC, Joo H, Maile MD, Aaronson KD, Burns ML, et al. Early detection of heart failure with reduced ejection fraction using perioperative data among noncardiac surgical patients: a machine-learning approach. *Anesth Analg*. (2020) 130:1188–200. doi: 10.1213/ANE.0000000000004630
11. Jalali A, Lonsdale H, Zamora LV, Ahumada L, Nguyen ATH, Rehman M, et al. Machine learning applied to registry data: development of a patient-specific prediction model for blood transfusion requirements during craniofacial surgery using the pediatric craniofacial perioperative registry dataset. *Anesth Analg*. (2021) 132:160–71. doi: 10.1213/ANE.0000000000004988
12. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care*. (2019) 23:112. doi: 10.1186/s13054-019-2411-z
13. Abujaber A, Fadlalla A, Gammoh D, Al-Thani H, El-Menyar A. Machine learning model to predict ventilator associated pneumonia in patients with traumatic brain injury: the C.5 decision tree approach. *Brain Inj*. (2021) 35:1095–102. doi: 10.1080/02699052.2021.1959060
14. Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevska O, AME Big-Data Clinical Trial Collaborative Group. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med*. (2019) 7:152. doi: 10.21037/atm.2019.03.29
15. van Niftrik CHB, van der Wouden F, Staartjes VE, Fierstra J, Stienen MN, Akeret K, et al. Machine learning algorithm identifies patients at high risk for early complications after intracranial tumor surgery: registry-based cohort study. *Neurosurgery*. (2019) 85:E756–64. doi: 10.1093/neuros/nyz145
16. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539
17. Wu Z, Liu Y, Xu J, Xie J, Zhang S, Huang L, et al. A ventilator-associated pneumonia prediction model in patients with acute respiratory distress syndrome. *Clin Infect Dis*. (2020) 71(Suppl. 4):S400–8. doi: 10.1093/cid/ciaa1518
18. Collins GS, Reitsma JB, Altman DG, Moons GK. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. (2015) 13:1. doi: 10.1186/s12916-014-0241-z
19. Bartoletti M, Giannella M, Scudeller L, Tedeschi S, Rinaldi M, Bussini L, et al. Development and validation of a prediction model for severe respiratory failure in hospitalized patients with SARS-CoV-2 infection: a multicentre cohort study (PREDI-CO study). *Clin Microbiol Infect*. (2020) 26:1545–53. doi: 10.2139/ssrn.3588558
20. Zhang Z. Variable selection with stepwise and best subset approaches. *Ann Transl Med*. (2016) 4:136. doi: 10.21037/atm.2016.03.35
21. Le S, Pellegrini E, Green-Saxena A, Summers C, Hoffman J, Calvert J, et al. Supervised machine learning for the early prediction of acute respiratory distress syndrome (ARDS). *J Crit Care*. (2020) 60:96–102. doi: 10.1016/j.jcrc.2020.07.019
22. Gannon WD, Lederer DJ, Biscotti M, Javaid A, Patel NM, Brodie D, et al. Outcomes and mortality prediction model of critically ill adults with acute respiratory failure and interstitial lung disease. *Chest*. (2018) 153:1387–95. doi: 10.1016/j.chest.2018.01.006
23. Lichtenstein DA, Mezière AG. Relevance of lung ultrasound in the diagnosis of acute respiratory failure: the BLUE protocol. *Chest*. (2008) 134:117–25. doi: 10.1378/chest.07-2800
24. Feng JF, Zhao X, Gurkoff GG, Van KC, Shahlaie K, Lyeth GB. Post-traumatic hypoxia exacerbates neuronal cell death in the hippocampus. *J Neurotrauma*. (2012) 29:1167–79. doi: 10.1089/neu.2011.1867
25. Okonkwo DO, Shutter LA, Moore C, Temkin NR, Puccio AM, Madden CJ, et al. Brain Oxygen optimization in severe traumatic brain injury phase-II: a phase II randomized trial. *Crit Care Med*. (2017) 45:1907–14. doi: 10.1097/CCM.0000000000002619
26. Ross BJ, Barker DE, Russell WL, Burns PR. Prediction of long-term ventilatory support in trauma patients. *Am Surg*. (1996) 62:19–25.
27. Brandenburg R, Brinkman S, de Keizer NF, Kesecioglu J, Meulenbelt J, de Lange WD. The need for ICU admission in intoxicated patients: a prediction model. *Clin Toxicol*. (2017) 55:4–11. doi: 10.1080/15563650.2016.1222616
28. Çiftci F, Çiledag A, Erol S, Kaya A. Non-invasive ventilation for acute hypercapnic respiratory failure in older patients. *Wien Klin Wochenschr*. (2017) 129:680–6. doi: 10.1007/s00508-017-1182-2
29. Carabias CS, Gomez PA, Panero I, Eiriz C, Castaño-León AM, Egea J, et al. Chitinase-3-like protein 1, serum amyloid A1, C-reactive protein, and procalcitonin are promising biomarkers for intracranial severity assessment of traumatic brain injury: relationship with glasgow coma scale and computed tomography volumetry. *World Neurosurg*. (2020) 134:e120–43. doi: 10.1016/j.wneu.2019.09.143
30. Poignant S, Vigué B, Balram P, Biais M, Carillon R, Cottenceau V, et al. A one-day prospective national observational study on sedation-analgesia of patients with brain injury in French Intensive Care Units: the SEDA-BIP-ICU (Sedation-Analgesia in Brain Injury Patient in ICU) Study. *Neurocrit Care*. (2021). doi: 10.1007/s12028-021-01298-x. [Epub ahead of print].
31. Decavèle M, Dreyfus A, Gatulle N, Weiss N, Houillier C, Demeret S, Mayaux et al. Clinical features and outcome of patients with primary central nervous system lymphoma admitted to the intensive care unit: a French national expert center experience. *J Neurol*. (2021) 268:2141–50. doi: 10.1007/s00415-021-10396-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 24 December 2021
doi: 10.3389/fmed.2021.786414

Efficacy and Safety of Anticoagulation Treatment in COVID-19 Patient Subgroups Identified by Clinical-Based Stratification and Unsupervised Machine Learning: A Matched Cohort Study

Yi Bian[1,2†], Yue Le[1,2†], Han Du[3], Junfang Chen[4], Ping Zhang[5], Zhigang He[1,2], Ye Wang[1,2], Shanshan Yu[1,2], Yu Fang[1,2], Gang Yu[1,2], Jianmin Ling[1,2], Yikuan Feng[1,2], Sheng Wei[6], Jiao Huang[7], Liuniu Xiao[1,2], Yingfang Zheng[1,2], Zhen Yu[8] and Shusheng Li[1,2]*

[1] Department of Emergency Medicine, Tongji Medical College, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, China, [2] Department of Critical Care Medicine, Tongji Medical College, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, China, [3] Germany Research Center for Artificial Intelligence, Saarland Informatics Campus, Saarbrücken, Germany, [4] Intelligent Medicine Research Center, Greater Bay Area Institute of Precision Medicine (Guangzhou), Fudan University, Guangzhou, China, [5] Department of Neurology, Tongji Medical College, Tongji Hospital, Huazhong University of Science and Technology, Wuhan, China, [6] Ministry of Education Key Laboratory of Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [7] Center for Evidence-Based and Translational Medicine, Zhongnan Hospital of Wuhan, Wuhan, China, [8] Department of Critical Care Medicine, Renmin Hospital of Wuhan University, Wuhan, China

Objective: To explore the efficacy of anticoagulation in improving outcomes and safety of Coronavirus disease 2019 (COVID-19) patients in subgroups identified by clinical-based stratification and unsupervised machine learning.

Methods: This single-center retrospective cohort study unselectively reviewed 2,272 patients with COVID-19 admitted to the Tongji Hospital between Jan 25 and Mar 23, 2020. The association between AC treatment and outcomes was investigated in the propensity score (PS) matched cohort and the full cohort by inverse probability of treatment weighting (IPTW) analysis. Subgroup analysis, identified by clinical-based stratification or unsupervised machine learning, was used to identify sub-phenotypes with meaningful clinical features and the target patients benefiting most from AC.

Results: AC treatment was associated with lower in-hospital death risk either in the PS matched cohort or by IPTW analysis in the full cohort. A higher incidence of clinically relevant non-major bleeding (CRNMB) was observed in the AC group, but not major bleeding. Clinical subgroup analysis showed that, at admission, severe cases of COVID-19 clinical classification, mild acute respiratory distress syndrome (ARDS) cases, and patients with a D-dimer level $\geq$0.5 μg/mL, may benefit from AC. During the hospital stay, critical cases and severe ARDS cases may benefit from AC. Unsupervised machine learning analysis established a four-class clustering model. Clusters 1 and

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Qilin Yang,
The Second Affiliated Hospital of Guangzhou Medical University, China
Chen Hui,
The First Affiliated Hospital of Soochow University, China
Yanfei Shen,
Zhejiang Hospital, China

***Correspondence:**
Shusheng Li
lishusheng@hust.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *30 September 2021*
Accepted: *26 November 2021*
Published: *24 December 2021*

Citation:
Bian Y, Le Y, Du H, Chen J, Zhang P, He Z, Wang Y, Yu S, Fang Y, Yu G, Ling J, Feng Y, Wei S, Huang J, Xiao L, Zheng Y, Yu Z and Li S (2021) Efficacy and Safety of Anticoagulation Treatment in COVID-19 Patient Subgroups Identified by Clinical-Based Stratification and Unsupervised Machine Learning: A Matched Cohort Study. Front. Med. 8:786414. doi: 10.3389/fmed.2021.786414

2 were non-critical cases and might not benefit from AC, while clusters 3 and 4 were critical patients. Patients in cluster 3 might benefit from AC with no increase in bleeding events. While patients in cluster 4, who were characterized by multiple organ dysfunction (neurologic, circulation, coagulation, kidney and liver dysfunction) and elevated inflammation biomarkers, did not benefit from AC.

Conclusions: AC treatment was associated with lower in-hospital death risk, especially in critically ill COVID-19 patients. Unsupervised learning analysis revealed that the most critically ill patients with multiple organ dysfunction and excessive inflammation might not benefit from AC. More attention should be paid to bleeding events (especially CRNMB) when using AC.

Keywords: COVID-19, anticoagulation, outcomes, mortality, bleeding events, unsupervised machine learning

INTRODUCTION

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has developed into a pandemic disease and affected nearly every country in the world. There is no comprehensive and strong clinical evidence to support the efficacy of any drugs that specifically target the SARS-CoV-2 (1). Previous research has found that coagulopathy is very common in COVID-19 patients, and includes thrombosis and coagulation abnormalities and dysfunction such as an elevated D-dimer level and prolonged prothrombin time (PT), respectively (2). Autopsy histopathologic analysis has identified widespread thrombosis and microangiopathy in small vessels and capillaries of the lung (3–5), which are different from the pathologies observed in respiratory failure caused by other diseases (3, 6–8). Some scholars have therefore proposed anticoagulation (AC) treatment as an integral part of systemic therapy in the early stage of COVID-19 (9). Generally, retrospective studies have suggested that AC may decrease mortality in COVID-19 patients (9, 10), However, these conclusions are not completely reliable nor applicable to all COVID-19 patients due to limitations in methodology such as no prospective control or matching cohort (9–11), large heterogeneity in anticoagulant therapy (9–12), and a lack of subgroup analysis (10, 11, 13). A recently completed randomized controlled trial (RCT) found that, compared with usual-care thromboprophylaxis, an initial strategy of therapeutic-dose anticoagulation did not result in a higher probability of survival in critically ill COVID-19 patients (14). The conclusion of this RCT may be inconsistent with that of previous retrospective studies. At present, the recommendations for empiric systemic AC treatment currently differ between COVID-19 management guidelines (15–17), with some recommending using anticoagulant drugs preventively for patients who have no contraindications to AC and a significantly increased D-dimer level, while others recommend that all hospitalized adults with COVID-19 should receive pharmacologic thromboprophylaxis with low molecular weight heparin (LMWH) rather than unfractionated heparin (UFH).

We conducted a retrospective cohort study using a comprehensive database of COVID-19 patients to investigate whether AC treatment was protective and safe for COVID-19 patients. Innovative analyses using propensity score matching (PSM) and inverse probability of treatment weighting (IPTW) were performed to balance baseline covariates, variates related to AC treatment assignment and variates related to the outcome between patients with or without AC treatment. Further sensitivity analyses were carried out to explore the association between outcome and duration, dosage and type of AC treatment. The second aim of the study was to identify the patients who benefited most from AC treatment using subgroup analysis that involved stratifying the data according to the severity of the acute respiratory distress syndrome (ARDS) (18), COVID-19 clinical classification (17), and D-dimer levels. Taking into account the heterogeneity of the patients, clinically relevant patient subpopulations were identified by unsupervised machine learning algorithms. The effectiveness of AC treatment was verified further in identified clusters.

MATERIALS AND METHODS

Ethics and Registration

This retrospective cohort study was approved by the ethics committee of Tongji Medical College, Huazhong University of Science and Technology (No. 2020-S220). The clinical trial was registered and verified by the Chinese Clinical Trial Registry (ChiCTR2000039855).

Patient Population

This single-center retrospective cohort study was conducted in two designated branches for COVID-19 patients in Tongji Hospital, an academic hospital affiliated to Tongji Medical College, Huazhong University of Science and Technology in Wuhan, China. All patients with confirmed COVID-19 admitted consecutively to these two institutions between Jan 25 to Mar 23, 2020, were enrolled retrospectively in the study. Approval was obtained from the ethics committee at our institution that the patients did not need to provide informed consent for inclusion in the study. Patients were assigned to three groups, including one group of patients with systemic AC treatment for at least 7 days, one group of patients with systemic AC treatment for <7 days and one group of patients without AC treatment. The

medications administered and clinical outcomes were followed up to June 4, 2020, when these two branches for exclusive COVID-19 treatment were closed. All COVID-19 patients were diagnosed according to the World Health Organization interim guidelines (19) and the Diagnosis and Treatment Protocol for COVID-19 Patients (Trial Version 8) (17). The exclusion criteria for the study were younger than 18 years, pregnant, length of stay <24 h, insufficient medical information, a history of severe comorbidities requiring surgical operation including, but not limited to, multiple trauma, a severe infection that required debridement, amputation or laparotomy, and patients who were classified again as COVID positive after RNA for SARS-CoV-2 was detected following their discharge from hospital.

Anticoagulation Exposure

AC treatment was defined as receiving either UFH, LMWH, Fondaparinux sodium, Argatroban, or direct-acting oral anticoagulants (DOACs) (mainly Rivaroxaban). The initiation of AC treatment was decided by the bedside physicians. Possible reasons for AC treatment were extracted from electronic case files. Immortal time is a gap period between exposure and initiation of follow-up (20). We carried out a Cox proportional hazards model with a time-dependent manner for the drug exposure in this study.

Outcomes, Definitions, and Data Collection

The primary outcome of this study was in-hospital mortality. The safety endpoints included bleeding events and thrombocytopenia. Major bleeding was defined according to the International Society on Thrombosis and Haemostasis (ISTH) statement (21) as those that resulted in death, were life-threatening, caused chronic sequelae, or consumed major healthcare resources. Hemorrhage that did not fit the criteria for the ISTH definition of major bleeding but required medical intervention was classified as clinically relevant non-major bleeding (CRNMB) (22). Other bleeding events which did not meet the criteria of either major bleeding or CRNMB, including bloody sputum, positive fecal occult blood test/gastric occult blood test and microscopic hematuria, were reported separately. Thrombocytopenia was defined as a platelet count $<100 \times 10^9$/L (23).

The CURB-65 score (21, 24), ARDS (18, 25, 26), and quick sequential organ failure assessment (qSOFA) (27) were defined according to the literature, while the COVID-19 clinical classification was made according to the Diagnosis and Treatment Protocol for COVID-19 Patients (Trial Version 8) (17). The detailed definition of ARDS and COVID-19 clinical classification is shown in **Supplementary Table 1**.

All the characteristics and clinical information of the patients were obtained from electronic medical and nursing record systems. This data included age, gender, current smoking history, comorbidities, laboratory results at admission, CURB-65 score and qSOFA score at admission, ARDS classification (18) and COVID-19 clinical classification at admission and during the hospital stay, antiviral therapies and other treatments during hospitalization, the level of oxygen therapy at admission, and the most intense level of oxygen therapy during hospitalization. Variables with missing data >20% were excluded from this analysis. Multiple imputations were conducted to address the presence of missing values.

Unsupervised Clustering

For this work, we used the K-Medoids clustering algorithm to partition our data into subclasses in an unsupervised manner. The K-Medoids algorithm randomly selects K samples in the training data as the medoids. The remaining samples are assigned to each subclass based on the pairwise dissimilarities. The sample, which is more similar to the medoid, is assigned to the corresponding subclass. Next, the medoids are updated based on the new results of subgrouping. These two steps are iterated multiple times until there is no change in the assignments. In particular, we used the Partitioning Around Medoids (PAM), which is the most common implementation of K-Medoids. For the dissimilarity measure, we adopted the Manhattan distance because it has better performance than the Euclidean distance for the data containing both binary and category variables (28).

A total of 25 variables representing the patients' clinical characteristics were used as input features in the unsupervised learning method, which included demographic features, comorbidities, vital signs, biomarkers, and oxygen therapy types at admission and during the hospital stay. Each patient in our database was presented as a vector with 25 dimensions. To prepare each patient's case as a vector for modeling training, we converted the binary variable as (0, 1), with the category variable represented by the corresponding categorical index. Normalization of the numerical variables was performed. After the normalization, we made sure that each numerical variable had a normal distribution.

Two prominent probabilistic model selection methods: Bayesian information criterion (BIC) and Akaike information criterion (AIC) were used to determine the optimal number of clusters in this work. In general, the measurement of AIC and BIC scores are similar. BIC penalizes the complexity of the model more than AIC (29).

Statistical Analysis

To minimize bias caused by the non-random allocation of potentially confounding covariates between the AC and non-AC groups, we adopted PSM methods (30). Propensity score (PS) was calculated using a logistic regression model, adjusted for the following covariates: level of oxygen therapy, clinical classification, high-sensitivity C reactive protein (hs-CRP), D-dimer levels, platelets count, CURB-65 score for the severity of pneumonia (31) at hospital admission, and the highest level of oxygen therapy during hospitalization. The match ratio was set at 1 to 3 and the maximum allowable distance (caliper) at 0.1 (32). To detect the selective bias potentially caused by this PSM, inverse probability of treatment weighting analysis (IPTW) was carried out based on the same variates used in PSM modeling (33).

Continuous variables were expressed as medians and interquartile ranges (IQR) and compared using the Mann-Whitney U test or Kruskal-Wallis H test. Categorical variables were compared using the Pearson $\chi 2$ test, continuity correction,

or Fisher's exact test, as appropriate. Differences between clusters identified by unsupervised machine learning were judged by two-tailed Bonferroni correction *post hoc* tests, with a *P*-value < (0.05/6) considered statistically significant. A Kaplan-Meier curve was used to analyze survival during hospitalization, with the data stratified according to AC treatment and PAM clustering subphenotypes.

Univariate and multivariate Cox proportional hazards regression was used to determine the risk factors for in-hospital mortality in the PS matched cohort. Residual imbalanced variates were included in the multivariate Cox proportional hazards regression for the PSM cohort. Cox regression analysis with IPTW adjusted covariates of important demographic characteristics as well as variates associated with outcomes either reported previously (34–36) or by general clinic consensus, which included age, gender, platelets count, PT, D-dimer, total bilirubin, lactate dehydrogenase, urea and hs-CRP. Competing risk model analyses were carried out using Fine-Gray tests, which considered death as a competing event for the safety endpoints, including bleeding events and thrombocytopenia. Sensitivity analyses were carried out according to the AC exposure duration, type and dosage in the full cohort. In subgroups analysis among PS matched cohort, univariate and multivariate logistic regression analysis was used to explore the association of outcomes and AC treatment. Residual imbalanced variates were included in the multivariate logistic regression here. The interaction effect between AC treatment and subgroups was also analyzed by logistic regression.

SPSS version 26.0 software (IBM Corp., Armonk, New York, USA) and SAS version 9.4 (SAS Institute Inc. Cary, NC, USA) were used for the statistical analyses and PS matching. The Kaplan-Meier survival plot and forest plot were constructed using GraphPad Prism version 4.0 software (GraphPad Software Inc., La Jolla, CA, USA). All tests were two-tailed, with a $P < 0.05$ considered statistically significant.

RESULTS

Clinical Characteristics of the Patients at Presentation

Two thousand four hundred and sixty nine confirmed COVID-19 patients admitted to Tongji Hospital between Jan 25 and Mar 23, 2020, were consecutively and unselectively reviewed. After excluding 197 patients who did meet study exclusion criteria, a total of 2,272 patients were identified for IPTW analysis and sensitivity analyses. In PSM modeling, 78 patients who received AC treatment for <7 days were not included for matching. Finally, PS matching yielded 165 patients in the AC group (patients who received AC for 7 days or longer) and 393 in the non-AC group (patients who did not receive AC) (**Figure 1**). Detailed AC treatment type, dosage, duration, time of initiation from admission and possible reasons for AC treatment were shown in **Supplementary Figure 1**. In the PS matched cohort, compared to the non-AC group, patients in the AC group were older (69 years, interquartile range [IQR] 60–78 vs. 65 years, IQR 53–71 years, $P < 0.001$) and had more comorbidities at admission (75.2% vs. 58.5%, $P < 0.001$, **Table 1**).

Primary and Secondary Outcomes

In the PS matched cohort, univariate Cox proportional hazard regression analysis showed a significantly lower probability of in-hospital death in the AC treatment group compared to that in the non-AC treatment group (hazard ratio [HR] = 0.450; 95% confidence interval [CI], 0.278 to 0.727; $P < 0.001$). Since there were still residual imbalances between AC and non-AC groups, multivariate Cox proportional hazard regression was carried out by adjusting imbalance covariates, including age, smoking, comorbidities, white blood cells, lactate dehydrogenase, urea, D-dimer, antiviral therapy, intravenous immunoglobulin, and oxygen therapy during hospitalization. Multivariate Cox analysis showed that AC treatment was associated with a lower probability of in-hospital mortality (adjusted HR = 0.249, 95% CI 0.143 to 0.436, $P < 0.001$, **Table 2**). To explore the immortal time bias, a further Cox proportional hazards model with a time-dependent manner for AC exposure was carried out in this study. This showed that the AC treatment was still associated with a lower probability of in-hospital death risk in the PS matched cohort (adjusted HR = 0.531, 95% CI 0.301 to 0.935, $P = 0.028$, **Figure 2**).

To detect the selective bias caused by PSM, IPTW analysis was performed in the full cohort including 2,272 patients. Multivariate Cox analysis again showed that AC treatment was associated with a lower probability of in-hospital mortality by comparing patients without AC treatment and patients with AC treatment for 7 days or longer (adjusted HR = 0.164; 95% CI: 0.104 to 0.260, $P < 0.001$). Sensitivity analyses were carried out according to the AC exposure duration, type and dosage in the full cohort. Duration of AC therapy <7 days was not associated with lower in-hospital mortality (adjusted HR = 1.018; 95% CI: 0.742 to 1.399, $P = 0.910$). In addition, AC treatment in all various dosages and types remained consistently associated with lower mortality (**Table 3**).

Secondary outcomes included bleeding events and thrombocytopenia. By adjusting baseline covariates and time-dependent AC exposure, it was revealed that AC treatment was associated with higher risk of total bleeding events (adjusted HR = 3.187, 95% CI, 1.846 to 5.504, $P < 0.001$), CRNMB (adjusted HR = 3.713 95% CI, 1.446 to 9.532, $P = 0.006$) and microscopic hematuria (adjusted HR = 2.624, 95% CI, 1.280 to 5.380, $P = 0.008$), but not associated with major bleeding ($P = 1.000$) and thrombocytopenia (adjusted HR = 2.167, 95% CI, 0.750 to 6.263, $P = 0.153$) (**Table 2**). Considering death as a competing event, a competing risk model analysis using the Fine-Gray test was conducted to explore the association between AC and bleeding events or thrombocytopenia. Finally, the AC group had no significantly higher risk of bleeding events (Fine-Gray test, $P = 0.500$) or thrombocytopenia (Fine-Gray test, $P = 0.911$) than the non-AC group when using death as a competing event in the model.

Clinical Subgroup Stratification

In-hospital mortality between the AC and the non-AC groups was compared in individuals stratified according to ARDS

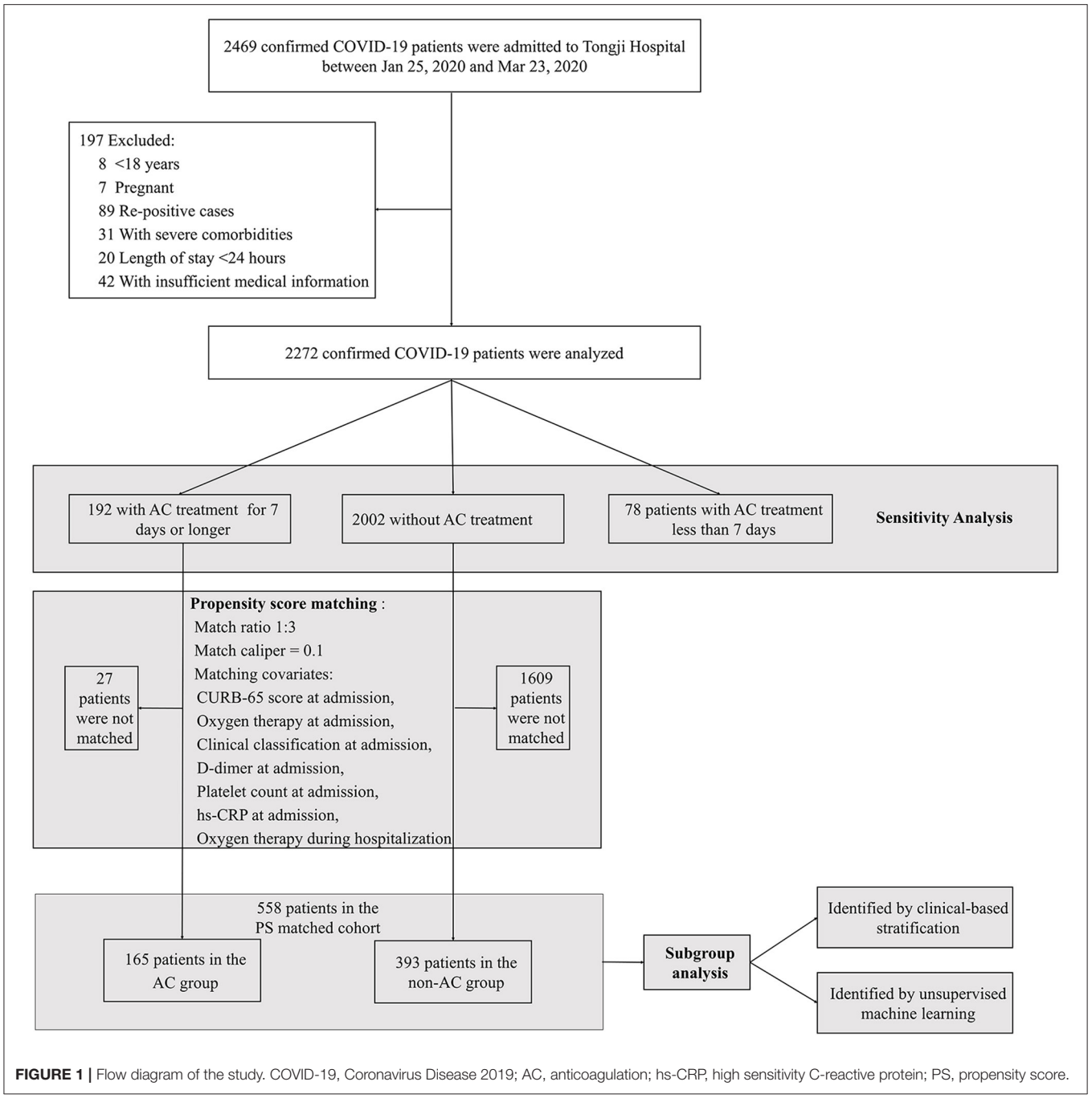


FIGURE 1 | Flow diagram of the study. COVID-19, Coronavirus Disease 2019; AC, anticoagulation; hs-CRP, high sensitivity C-reactive protein; PS, propensity score.

classification, COVID-19 clinical classification, and D-dimer levels at both hospital admission and during hospitalization. At hospital admission, AC treatment was associated with lower in-hospital mortality in subgroups of mild ARDS (adjusted odds ratio [OR] = 0.005, 95% CI, 0.000–0.174, $P = 0.004$), severe COVID-19 cases (adjusted OR = 0.076, 95% CI, 0.024–0.236, $P < 0.001$) and patients with a D-dimer level $\geq 0.5\,\mu g/mL$ (adjusted OR = 0.042, 95% CI, 0.003–0.603, $P = 0.020$). During the hospital stay, AC treatment was associated with lower in-hospital mortality among patients who developed to severe ARDS (adjusted OR = 0.046, 95% CI, 0.013–0.157, $P < 0.001$) or critical COVID-19 (adjusted OR = 0.095, 95% CI, 0.034–0.266, $P < 0.001$) (**Figure 3**).

The interaction effect between AC treatment and the subgroups was also analyzed by logistic regression (**Supplementary Table 2**). An interaction effect was observed between the subgroups identified by ARDS at admission ($P < 0.001$) or clinical classification at admission ($P < 0.001$) and AC treatment.

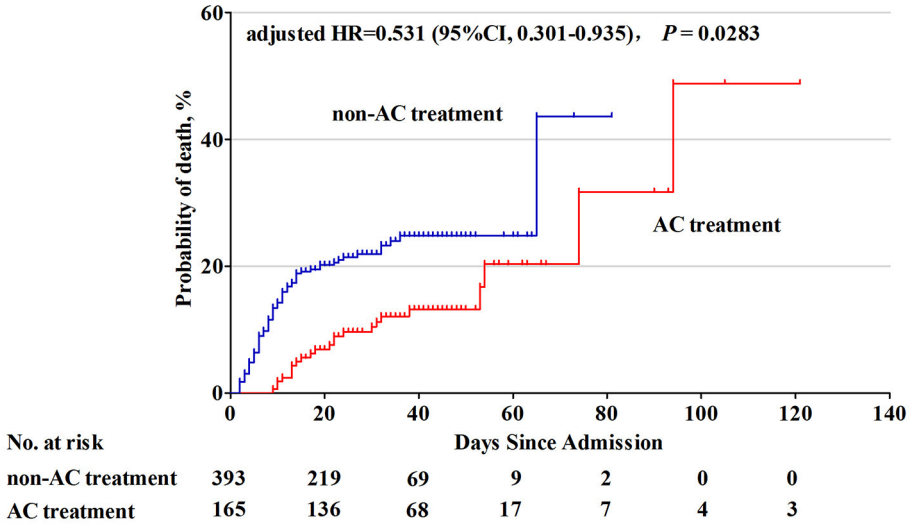


FIGURE 2 | Cumulative probability of death in COVID-19 patients with and without AC treatment. Patients with AC treatment had a lower probability of in-hospital mortality than those without AC treatment (adjusted* HR = 0.531, 95% CI: 0.301–0.935, $P = 0.0283$). *Adjusted for time-dependent AC exposure and baseline covariates including age, smoking, comorbidities, white blood cells, D-dimer, lactate dehydrogenase, urea, antiviral therapy, intravenous immunoglobulin, convalescent plasma therapy, and oxygen therapy in hospitalization. HR, hazard ratio; CI, confidence interval; AC, anticoagulation.

PAM Clustering Analysis

The Optimal Number of Clustering Determination

BIC and AIC scores were used to automatically select the best number of clusters K for our model. As shown in **Figure 4**, we tried to find the optimal K by conducting an exhaustive search of the possible K values. By and large, the results from AIC and BIC are proportional. BIC score suggests that four clusters are optimal. For the AIC score, a shape "elbow point" is also indicated at four, after four the decrease is becoming notably smaller. Even though a better fitness of the data might be achieved by increasing the number of clusters, an additional cost will be needed including the over-fitting issue and the complexity of interpreting clinically plausible subgroups. Therefore, four clusters were selected to be optimal in this work.

Clinical Features of Patients' Subgroups

The clinical features of the four clusters are shown in **Tables 4**, **5** and **Figure 5**. Patients in clusters 1 and 2 were non-critical cases. Patients in clusters 1 and 2 had no ARDS at admission then developed mild ARDS during the hospital stay. They mainly needed nasal cannula oxygenation at admission as well as during hospitalization. Within these non-critical clusters, cluster 1 was characterized by the youngest age (51 years, IQR 39–63) and had the least number of comorbidities (0, IQR 0–1). Compared to clusters 1 and 2, clusters 3 and 4 were critical cases. In laboratory testing results at admission, both clusters 3 and 4 had significantly higher neutrophils count, lower lymphocytes and platelets count, higher lactate dehydrogenase and hs-CRP, higher urea and creatinine, elevated prothrombin time and D-dimer. In vital signs, clusters 3 and 4 had significantly lower peripheral blood oxygen saturation and respiratory rate at admission. Either at admission or during the hospital stay, both clusters 3 and 4 had significantly more severe levels of ARDS and need a higher level of oxygen therapy. Notably, cluster 4 exhibited as the most critical sub-phenotype. Compared to cluster 3, patients in cluster 4 had significantly excessive inflammation (elevated white blood cells, neutrophils and lactate dehydrogenase), organ dysfunction (higher total bilirubin and urea), severe coagulopathy (elevated prothrombin time and D-dimer), unstable hemodynamics (higher rate of vasopressor use) and neurologic dysfunction (disturbance of consciousness). Compared to the other three clusters, cluster 4 had severe ARDS either at admission or during the hospital stay. As a result, cluster 4 also needed the highest level of oxygen therapy accordingly.

AC Treatment and Outcomes in Different Patients Clusters

In the non-critical cluster (clusters 1 and 2), we found no significantly lower in-hospital death risk associated with AC treatment (**Figure 6**). In cluster 3, patients who received AC treatment had a significantly lower in-hospital mortality than those who did not receive this treatment (adjusted OR = 0.027, 95% CI, 0.005 to 0.134, $P < 0.001$). However, patients in cluster 4, who had elevated inflammation biomarkers and even severe multi-organ dysfunction, did not benefit from AC treatment. For safety endpoint, AC treatment was not associated with increasing bleeding events (**Figure 6**). In addition, the interaction effect was found between the AC treatment and sub-phenotypes identified by PAM clustering ($P < 0.001$, **Supplementary Table 2**).

DISCUSSION

COVID-19 infections have affected patients globally. The ISTH pointed out that COVID-19 patients develop a clinically

TABLE 1 | Patients Baseline Characteristics and Treatments in propensity score matched cohort.

	No. (%)			
	Total (*n* = 558)	**AC treatment (*n* = 165)**	**Non-AC treatment (*n* = 393)**	***P*-value**
Age, median (IQR), years	66 (56, 73)	69 (60, 78)	65 (53, 71)	<0.001
Gender				
Male	307 (55.0)	91 (55.2)	216 (55.0)	0.967
Female	251 (45.0)	74 (44.8)	177 (450)	
Current smoking	40 (7.2)	18 (10.9)	22 (5.6)	0.026
Comorbidities	354 (63.4)	124 (75.2)	230 (58.5)	<0.001
Diabetes	118 (21.1)	41 (24.8)	77 (19.6)	0.165
Hypertension	247 (44.3)	90 (54.5)	157 (39.9)	0.002
Cardiovascular disease	71 (12.7)	31 (18.8)	40 (10.2)	0.005
CODP	45 (8.1)	20 (12.1)	25 (6.4)	0.023
Chronic kidney disease	19 (3.4)	6 (3.6)	13 (3.3)	0.845
Chronic liver disease	18 (3.2)	4 (2.4)	14 (3.6)	0.487
Autoimmune disease	8 (1.4)	2 (1.2)	6 (1.5)	1.000
Immunosuppression	2 (0.4)	0 (0)	2 (0.5)	1.000
Malignancy	21 (3.8)	3 (1.8)	18 (4.6)	0.118
Oxygen therapy at admission				
Without oxygen inhalation	100 (17.9)	25 (15.2)	75 (19.1)	0.206
Nasal cannula	375 (67.2)	110 (66.7)	265 (67.4)	
Face mask with reservivor bag	46 (8.2)	20 (12.1)	26 (6.6)	
High-flow nasal cannula	4 (0.7)	2 (1.2)	2 (0.5)	
Non-invasive ventilation(bi-level)	20 (3.6)	4 (2.4)	16 (4.1)	
Invasive mechanical ventilation	13 (2.3)	4 (2.4)	9 (2.3)	
ARDS at admission				
No ARDS	331 (59.3)	85 (51.5)	246 (62.6)	0.067
Mild	120 (21.5)	45 (27.3)	75 (19.1)	
Moderate	71 (12.7)	25 (15.2)	46 (11.7)	
Severe	36 (6.5)	10 (6.1)	26 (6.6)	
Clinical classification at admission				
Moderate	62 (11.1)	18 (10.9)	44 (11.2)	0.537
Severe	459 (82.3)	139 (84.2)	320 (81.4)	
Critical	37 (6.6)	8 (4.8)	29 (7.4)	
CURB-65 score at admission	1 (0, 2)	1 (0, 2)	1 (0, 1)	0.099
qSOFA score at admission	0 (0, 1)	0 (0, 1)	0 (0, 1)	0.608
Initial laboratory parameters, median (IQR)				
White blood cells, $\times 10^9$/L	6.29 (4.89, 8.58)	7.15 (5.40, 9.57)	5.97 (4.60, 8.02)	<0.001
Neutrophils, $\times 10^9$/L	4.71 (3.12, 7.04)	5.65 (4.06, 7.78)	4.24 (2.90,6.37)	<0.001
Lymphocytes, $\times 10^9$/L	0.90 (0.60, 1.31)	0.84 (0.60, 1.15)	0.94 (0.61, 1.34)	0.060
Platelets, $\times 10^9$/L	217.0 (153.0, 291.3)	209.0 (143.5, 278.0)	219.0 (159.0, 293.0)	0.129
Total bilirubin, mmol/L	9.6 (7.2, 13.7)	10.7 (7.0, 14.8)	9.4 (7.2, 13.1)	0.072
Lactate dehydrogenase, U/L	321.0 (237.0, 450.3)	357.0 (262.5, 468.0)	313.0 (228.0, 434.5)	0.002
Urea, mmol/L	5.0 (3.6, 7.0)	5.2 (4.1, 7.4)	4.8 (3.4, 6.8)	0.019
hs-CRP, mg/L	43.6 (13.0, 102.0)	50.9 (1.0, 112.7)	36.9 (10.6, 99.0)	0.009
Prothrombin time, s	14.0 (13.4, 15.0)	14.1 (13.5, 15.2)	14.0 (13.4, 14.9)	0.142
D-dimer, mg/mL	1.36 (0.58, 2.98)	2.22 (1.11, 6.21)	1.11 (0.56, 2.58)	<0.001
Creatine	72 (58, 89)	74 (59, 90)	71 (58, 89)	0.707
Antiviral therapy	522 (93.5)	162 (98.2)	360 (91.6)	0.004
Other treatments				
Intravenous immunoglobulin	186 (33.3)	82 (49.7)	104 (26.5)	<0.001
Corticosteroid	276 (49.5)	107 (64.8)	169 (43.0)	<0.001
Convalescent plasma	22 (3.9)	15 (9.1)	7 (1.8)	<0.001

(Continued)

TABLE 1 | Continued

	No. (%)			
	Total (n = 558)	**AC treatment (n = 165)**	**Non-AC treatment (n = 393)**	***P*-value**
Oxygen therapy in hospitalization				
Without oxygen inhalation	6 (1.1)	1 (0.6)	5 (1.3)	<0.001
Nasal cannula	328 (58.8)	85 (51.5)	243 (61.8)	
Face mask with reservivor bag	43 (7.7)	15 (9.1)	28 (7.1)	
High-flow nasal cannula	25 (4.5)	8 (4.8)	17 (4.3)	
Non-invasive ventilation(bi-level)	79 (14.2)	15 (9.1)	64 (16.3)	
Invasive mechanical ventilation	72 (12.9)	36 (21.8)	36 (9.2)	
ECMO	5 (0.9)	5 (3.0)	0 (0)	

Variables represented the poorest value of the first day at admission. AC, anticoagulation; IQR, interquartile range; COPD, Chronic obstructive pulmonary disease; ARDS, acute respiratory distress syndrome; qSOFA, quick sequential organ failure assessment; hs-CRP, high sensitive C reacting protein; ECMO, extracorporeal membrane oxygenation.

TABLE 2 | Primary and secondary outcomes of PS matched cohort.

	No. (%)					
	Total (n = 558)	**AC group (n = 165)**	**Non-AC group (n = 393)**	**Crude HR (95% CI)**	**Adjusted* HR (95% CI)**	**Adjusted* HR for time-dependent AC exposure (95% CI)**
Primary outcomes						
In-hospital mortality	107 (19.2)	23 (13.9)	84 (21.4)	0.450 (0.278–0.727)	0.249 (0.143–0.436)	0.531 (0.301–0.935)
Secondary outcomes						
Bleeding events	121 (21.7)	42 (25.5)	79 (20.1)	0.673 (0.460–0.984)	0.675 (0.413–1.104)	3.187 (1.846–5.504)
Major bleeding	4 (0.7)	0 (0)	4 (1.0)	0.017 (0–58.149)	–	–
CRNMB	32 (5.7)	17 (10.3)	15 (3.8)	1.339 (0.663–2.705)	1.149 (0.465–2.839)	3.713 (1.446–9.532)
Bloody sputum	9 (1.6)	3 (1.8)	6 (1.5)	0.628 (0.155–2.544)	0.201 (0.008–4.896)	0.958 (0.051–18.063)
Positive FOBT/GOBT	5 (0.9)	4 (2.4)	1 (0.3)	4.011 (0.447–35.976)	–	–
Microscopic hematuria	80 (14.3)	25 (15.2)	55 (14.0)	0.597 (0.370–0.965)	0.443 (0.232–0.846)	2.624 (1.280–5.380)
Thrombocytopenia	79 (14.2)	25 (15.2)	54 (13.8)	0.608 (0.372–0.993)	0.665 (0.334–1.327)	2.167 (0.750–6.263)

**Adjusted for baseline covariates including age, smoking, comorbidities, white blood cells, D-dimer, lactate dehydrogenase, urea, antiviral therapy, intravenous immunoglobulin, convalescent plasma therapy, and oxygen therapy in hospitalization. PS, propensity score; AC, anticoagulation; HR, hazard ratio; CI, confidence interval; CRNMB, clinically relevant non-major bleeding; FOBT, fecal occult blood test; GOBT, gastric occult blood test.*

TABLE 3 | Hazard ratio for in-hospital mortality in the full cohort by AC treatment duration, dosage and type in sensitivity analyses.

	No. of in-hospital death/total no. (%)	**Crude HR (95% CI)**	**Adjusted* HR (95% CI)**	**Adjusted* HR for IPTW model[#] (95% CI)**
Non-AC treatment	91/2002 (4.5)	Reference	Reference	Reference
Duration				
AC treatment <7 days	44/78 (56.4)	13.600 (9.482–19.507)	3.424 (2.242–5.230)	1.018 (0.742–1.399)
AC treatment for 7 days or longer	38/192 (19.8)	2.881 (1.943–4.272)	0.864 (0.560–1.335)	0.164 (0.104–0.260)
Dosage				
Low dose thromboprophylaxis	43/176 (24.4)	4.354 (3.020–6.277)	1.387 (0.923–2.085)	0.498 (0.329–0.754)
Intermediate dose thromboprophylaxis	21/55 (38.2)	6.702 (4.153–10.816)	1.492 (0.870–2.560)	0.349 (0.224–0.545)
Therapeutic dose anticoagulation	18/39 (46.2)	6.959 (4.131–11.723)	1.575 (0.905–2.742)	0.225 (0.132–0.384)
Type				
LMWH	80/262 (30.5)	5.120 (3.878–7.162)	1.432 (1.004–2.043)	0.382 (0.271–0.537)
Non-LMWH	2/8 (25.0)	3.590 (0.866–14.884)	2.029 (0.482–8.538)	0.094 (0.019–0.466)

**Adjusted for baseline covariates including age, gender, levels of platelets count, prothrombin time, D-dimer, total bilirubin, lactate dehydrogenase, urea, and high-sensitivity C reactive protein.*

[#]Covariates in IPTW model: level of oxygen therapy, clinical classification, high-sensitivity C reactive protein and D-dimer levels, Platelet count at admission, CURB-65 score at hospital admission, and the highest level of oxygen therapy during hospitalization.

AC, anticoagulation; HR, hazard ratio; CI, confidence interval; IPTW, inverse probability of treatment weighting analysis; LMWH, low molecular weight heparin.

Subgroup	AC *no. of events/total no. (%)*	Non-AC *no. of events/total no. (%)*	Crude Odds Ratio(95%CI)
ARDS at admission			
No ARDS	8/85 (9.4)	17/246 (6.9)	1.400 (0.581-3.371)
Mild	9/45 (20.0)	20/75 (26.7)	0.688 (0.282-1.677)
Moderate	2/25 (8.0)	25/46 (54.3)	0.073 (0.015-0.347)
Severe	4/10 (40.0)	22/26 (84.6)	0.121 (0.023-0.634)
ARDS in hospitalization			
No ARDS	0/21 (0)	0/126 (0)	-
Mild	0/53 (0)	0/114 (0)	-
Moderate	0/27 (0)	2/43 (4.7)	-
Severe	23/64 (35.9)	82/110 (74.5)	0.192 (0.098-0.373)
Clinical classification at admission			
Moderate	3/18 (16.7)	1/44 (2.3)	8.600 (0.830-89.122)
Severe	18/139 (12.9)	58/320 (18.1)	0.672 (0.380-1.189)
Critical	2/8 (25.0)	25/29 (86.2)	0.053 (0.008-0.363)
Clinical classification in hospitalization			
Moderate	0/1 (0)	0/4 (0)	-
Severe	0/103 (0)	1/283 (0.4)	-
Critical	23/61 (37.7)	83/106 (78.3)	0.168 (0.084-0.336)
D-dimer at admission, mg/mL			
D-dimer<0.5	0/8 (0)	5/86 (5.8)	-
D-dimer ⩾0.5	23/157 (14.6)	79/307 (25.7)	0.495 (0.297-0.826)
0.5⩽D-dimer<2.5	9/80 (11.3)	40/204 (19.6)	0.520 (0.239-1.128)
2.5⩽D-dimer<5.0	4/33 (12.1)	15/46 (32.6)	0.285 (0.085-0.959)
D-dimer ⩾5.0	10/44 (22.7)	24/57 (42.1)	0.404 (0.168-0.975)
Over all	23/165 (13.9)	84/393 (21.4)	0.596 (0.361-0.985)

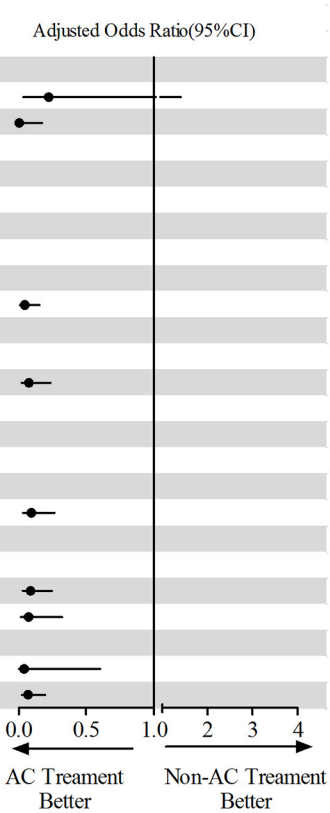


FIGURE 3 | Subgroup analysis based on clinical stratification of in-hospital mortality between AC treatment and non-AC treatment patients. The multivariate logistic regression analysis adjusted for baseline covariates including age, smoking, comorbidities, white blood cells, D-dimer, lactate dehydrogenase, urea, antiviral therapy, intravenous immunoglobulin, convalescent plasma therapy, and oxygen therapy in hospitalization. ARDS, acute respiratory distress syndrome; AC, anticoagulation.

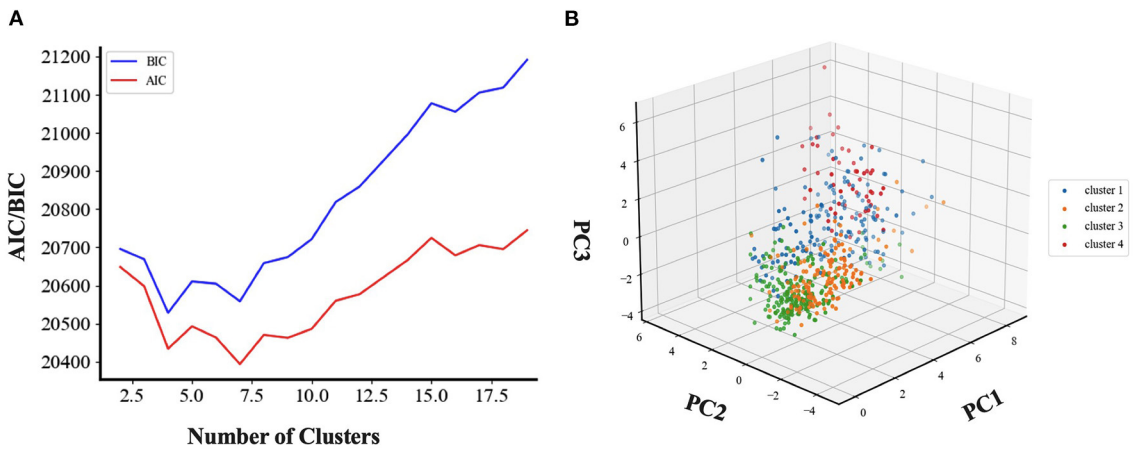


FIGURE 4 | Model selection and cluster visualization. The selection of the best number of clusters K is based on BIC and AIC score. K = 4 was chosen after comparing the BIC score of models with different number of clusters by unsupervised clustering analysis. For AIC score, K = 4 is also a good choice for the trade-off between model complexity and the fitting of the data. **(A)** Unsupervised clustering analysis for choosing the best number of clusters. BIC, Bayesian information criteria; AIC, Akaike information criterion. **(B)** Three-dimensional visualization of clustering results. We visualize our clustering results in three-dimensional space. The high dimensional training data was projected into three-dimensional space by using principal component analysis. PC, Principal component.

TABLE 4 | Variables included in the PAM-based clustering model.

	Cluster 1 (*n* = 144)	Cluster 2 (*n* = 203)	Cluster 3 (*n* = 158)	Cluster 4 (*n* = 53)	*P*-value
Age	51 (39, 63) abc	69 (62, 76)	67 (59, 74)	68 (63, 72)	<0.001
Gender	a	de			
Female	42 (29.2, −4.4)	156 (76.8, 11.4)	38 (24.1, −6.2)	15 (23.8, −2.6)	<0.001
Male	102 (70.8, 4.4)	47 (23.2, −11.4)	120 (75.9, 6.2)	38 (71.7, 2.6)	
No. of comorbidities	0 (0,1) abc	1 (0, 2)	1 (0, 2)	1 (1, 2)	<0.001
Initial laboratory parameters					
White blood cells, ×109/L	5.62 (4.47, 7.81) c	5.97 (4.89, 7.60) e	6.61 (4.72, 8.84) f	11.38 (8.96, 15.77)	<0.001
Neutrophils, ×109/L	3.79 (2.62, 5.44) bc	4.22 (3.03, 5.71) de	5.46 (3.47, 7.74) f	10.55 (7.64, 14.68)	<0.001
Lymphocytes, ×109/L	1.27 (0.90, 1.58) bc	1.06 (0.77, 1.42) de	0.61 (0.46, 0.82)	0.51 (0.36, 0.76)	<0.001
Platelets, ×109/L	283 (223, 342) abc	226 (177, 296) de	161 (127, 222)	157 (102, 216)	<0.001
Total bilirubin, mmol/L	8.9 (6.4, 12.1) bc	8.3 (6.4, 12.1) de	11.6 (8.3, 14.4) f	14.6 (10.6, 23.4)	<0.001
Lactate dehydrogenase, U/L	258 (205, 320) bc	284 (224, 344) de	444 (333, 524) f	601 (458, 795)	<0.001
Urea, mmol/L	3.95 (3.10, 5.28) bc	4.60 (3.30, 5.70) de	6.35 (4.70, 9.20) f	9.80 (6.60, 13.85)	<0.001
Creatinine, μmol/L	70 (58, 82) bc	63 (54, 78) de	83 (66, 101)	91 (75, 121)	<0.001
hs-CRP, mg/L	17.0 (6.7, 55.6) bc	17.0 (5.9, 43.3) de	109.8 (69.9, 162.3)	112.3 (71.4, 186.3)	<0.001
Prothrombin time, s	13.7 (13.2, 14.2) bc	13.8 (13.2, 14.3) de	14.5 (13.8, 15.4) f	16.2 (15.1, 17.6)	<0.001
D-dimer, mg/mL	0.91 (0.39, 2.10) bc	1.17 (0.58, 2.59) de	1.58 (0.86, 3.01) f	21.00 (2.72, 21.00)	<0.001
Hemoglobin, g/L	131 (120, 141) a	119 (109, 128) de	131 (120, 141)	132 (117, 142)	<0.001
At admission					
Oxygen therapy	bc	de	f		
Without oxygen inhalation	32 (22.2, 1.6)	47 (23.2, 2.4)	20 (12.7, −2.0)	1 (1.9, −3.2)	<0.001
Nasal cannula	105 (72.9, 1.7)	147 (72.4, 2.0)	103 (65.2, −0.6)	20 (37.7, −4.8)	
Face mask with reservoir bag	3 (2.1, −3.1)	8 (3.9, −2.8)	23 (14.6, 3.4)	12 (22.6, 4.0)	
High-flow nasal cannula	1 (0.7, 0)	0 (0, −1.5)	1 (0.6, −0.1)	2 (3.8, 2.8)	
NIV	2 (1.4, −1.6)	0 (0, −3.4)	6 (3.8, 0.2)	12 (22.6, 7.8)	
IMV	1 (0.7, −1.5)	1 (0.5, −2.2)	5 (3.2, 0.8)	6 (11.3, 4.6)	
Vasopressor	0 (0, −1.9) c	0 (0, −2.4) e	2 (1.3, −0.6) f	8 (15.1, 7.7)	<0.001
Disturbance of consciousness	0 (0, −3.0) c	2 (1.0, −2.9) e	6 (3.8, −0.4) f	16 (30.2, 9.8)	<0.001
SpO_2	97 (95, 98) bc	97 (95, 99) de	93 (86, 96)	90 (79, 96)	<0.001
Respiratory rate, /min	20 (20, 22) bc	20 (20, 22) de	20 (20, 25)	25 (20, 32)	<0.001
Heart rate, /min	92 (84, 105) a	86 (77, 99) d	96 (86, 109)	88 (78, 113)	<0.001
MAP, mmHg	97 (90, 104)	97 (89, 106)	96 (89, 107)	99 (91, 107)	0.771
During hospitalization					
Oxygen therapy	bc	de	f		
Without oxygen inhalation	4 (2.8, 2.3)	2 (1.0, −0.2)	0 (0, −1.5)	0 (0, −0.8)	<0.001
Nasal cannula	121 (84.0, 7.1)	173 (85.2, 9.6)	34 (21.5, −11.2)	0 (0, −9.1)	
Face mask with reservoir bag	8 (5.6, −1.1)	13 (6.4, −0.9)	22 (13.9, −0.9)	0 (0, −2.2)	
High-flow nasal cannula	4 (2.8, −1.1)	6 (3.0, −1.3)	14 (8.9, 3.1)	1 (1.9, −1.0)	
NIV	3 (2.1, −4.8)	1 (0.5, −7.0)	57 (36.1, 9.3)	18 (34.0, 4.3)	
IMV	3 (2.1, −4.5)	8 (3.9, −4.8)	27 (17.1, 1.9)	34 (64.2, 11.7)	
ECMO	1 (0., −0.3)	0 (0, −1.7)	4 (2.5, 2.6)	0 (0, −0.7)	
Vasopressor	3 (2.1, −5.7) bc	7 (3.4, −6.6) de	43 (27.2, 3.8) f	45 (84.9, 13.5)	<0.001
Lowest SpO_2	95 (93, 96) bc	95 (92, 96) de	89 (77, 94) f	67 (56, 80)	<0.001

Quantitative variables are expressed as medians (interquartile ranges). Categorical variables are expressed as No. (%, adjusted standardized residuals). It is considered that the difference between the actual frequency and the expected frequency of this value is statistically significant if the absolute value of the adjusted standardized residual is >2. Bonferroni correction for multiple comparison was used in the comparison of cluster analysis results. The letters "a" to "f" indicate a significant difference between two group, respectively (a = cluster 1 vs. cluster 2, b = cluster 1 vs. cluster 3, c = cluster 1 vs. cluster 4, d = cluster 2 vs. cluster 3, e = cluster 2 vs. cluster 4, f = cluster 3 vs. cluster 4. PAM, partitioning around medoids; hs-CRP, high sensitivity C-reactive protein; NIV, noninvasive ventilation (bi-level); IMV, invasive mechanical ventilation; SpO_2, peripheral blood oxygen saturation; MAP, mean arterial pressure; ECMO, extracorporeal membrane oxygenation.

TABLE 5 | Outcome and variables not included in the PAM-based clustering model.

	Cluster 1 (*n* = 144)	Cluster 2 (*n* = 203)	Cluster 3 (*n* = 158)	Cluster 4 (*n* = 53)	*P*-value
Symptoms					
Fever	120 (83.3, 0.4)	159 (78.3, −1.8)	135 (85.4, 1.2)	45 (84.9, 0.5)	0.306
Cough	109 (75.7, 0.1)	145 (71.4, −1.6)	123 (77.8, 0.9)	43 (81.1, 1.0)	0.367
Expectoration	66 (45.8, 0.3)	86 (42.4, −0.8)	75 (47.5, 0.8)	22 (41.5, −0.5)	0.746
Shortness of breath	74 (51.4, −1.7)	117 (57.6, 0.1)	100 (63.3, 1.8)	29 (54.7, −0.4)	0.210
Myalgia	27 (18.8, −0.8)	38 (18.7, −1.0)	41 (25.9, 1.8)	11 (20.8, 0)	0.332
Fatigue	48 (33.3, −0.3)	68 (33.5, −0.3)	57 (36.1, 0.5)	19 (35.8, 0.2)	0.942
Diarrhea	23 (16.0, −2.2)	50 (24.6, 0.9)	39 (24.7, 0.7)	14 (26.4, 0.7)	0.177
Nausea/vomiting	17 (11.8, −0.7)	29 (14.3, 0.4)	23 (14.6, 0.5)	6 (11.3, −0.5)	0.842
At admission					
FiO_2	0.30 (0.27, 0.30) bc	0.30 (0.27, 0.30) de	0.36 (0.30, 0.42) f	0.51 (0.38, 0.63)	<0.001
SBP, mmHg	124 (117, 136) ac	131 (121, 143)	131 (120, 146)	135 (122, 151)	0.004
DBP, mmHg	81 (76, 90)	79 (72, 87)	78 (73, 88)	80 (72, 88)	0.258
CURB-65 score	0 (0, 1) abc	1 (1, 1) de	1 (1, 2) f	2 (2, 3)	<0.001
ARDS at admission	bc	de	f		
No ARDS	108 (75.0, 4.4)	170 (83.7, 8.9)	49 (31.0, −8.6)	4 (7.5, −8.1)	<0.001
Mild	29 (20.1, −0.5)	26 (12.8, −3.8)	55 (34.8, 4.8)	10 (18.9, −0.5)	
Moderate	5 (3.5, −3.9)	6 (3.0, −5.2)	41 (25.9, 5.9)	19 (35.8, 5.3)	
Severe	2 (1.4, −2.9)	1 (0.5, −4.3)	13 (8.2, 1.1)	20 (37.7, 9.7)	
Initial laboratory parameters					
Hematocrit, %	38 (34, 40) a	35 (32, 38) de	37 (34, 40)	38 (34, 41)	<0.001
ALT, U/L	28 (19, 48) a	20 (13, 36) d	29 (17, 47)	24 (19, 41)	<0.001
AST, U/L	27 (20, 38) bc	23 (18, 34) de	40 (28, 62)	41 (26, 56)	<0.001
Total protein, g/L	69.6 (66.2, 73.5) bc	68.0 (64.6, 72.7) de	65.9 (62.5, 70.5)	63.3 (60.8, 69.9)	<0.001
Albumin, g/L	35.7 (31.5, 39.5) bc	34.4 (31.4, 38.3) de	31.4 (28.4, 34.2)	29.3 (26.7, 31.8)	<0.001
Globulins, g/L	33.3 (29.9, 37.8) b	33.5 (30.1, 37.1) d	34.9 (31.4, 39.1)	34.8 (31.5, 39.1)	0.003
During hospitalization					
Highest FiO_2	0.30 (0.30, 0.36) bc	0.33 (0.30, 0.36) de	0.66 (0.45, 1.00) f	1.00 (0.70, 1.00)	<0.001
ARDS during hospitalization	bc	de	f		
No ARDS	62 (43.1, 5.3)	83 (40.9, 5.9)	2 (1.3, −8.5)	0 (0, −4.6)	<0.001
Mild	60 (41.7, 3.6)	80 (39.4, 3.7)	27 (17.1, −4.2)	0 (0, −5.0)	
Moderate	11 (7.6, −2.1)	22 (10.8, −0.9)	35 (22.2, 4.3)	2 (3.8, −2.0)	
Severe	11 (7.6, −7.1)	18 (8.9, −8.6)	94 (59.5, 9.1)	51 (96.2, 10.7)	
In-hospital mortality	2 (1.4, −6.3) bc	6 (3.0, −7.4) de	52 (32.9, 5.2) f	47 (88.7, 13.5)	<0.001
Survival time, days	23 (17, 34) abc	31 (18, 40) e	31 (14, 44) f	10 (6, 17)	<0.001
Bleeding events	21 (14.6, −2.4) bc	31 (15.3, −2.8) de	51 (32.3, 3.8)	18 (34.0, 2.3)	<0.001
Major bleeding	0 (0, −1.2)	0 (0, −1.5)	4 (2.5, 3.2)	0 (0, −0.7)	0.021
CRNMB	6 (4.2, −0.9)	7 (3.4, −1.8)	14 (8.9, 2.0)	5 (9.4, 1.2)	0.070
Bloody sputum	2 (1.4, −0.2)	2 (1.0, −0.9)	5 (3.2, 1.8)	0 (0, −1.0)	0.378
Microscopic hematuria	12 (8.3, −2.4) bc	21 (10.3, −2.0) de	34 (21.5, 3.0)	13 (24.5, 2.2)	<0.001
Positive FOBT/GOBT	1 (0.7, −0.3)	2 (1.0, 0.2)	1 (0.6, −0.4)	1 (1.9, 0.8)	0.713
Thrombocytopenia	5 (3.5, −4.3) bc	7 (3.4, −5.5) de	37 (23.4, 3.9) f	30 (56.6, 9.3)	<0.001

Quantitative variables are expressed as medians (interquartile ranges). Categorical variables are expressed as No. (%, adjusted standardized residuals). It is considered that the difference between the actual frequency and the expected frequency of this value is statistically significant if the absolute value of the adjusted standardized residual is >2. Bonferroni correction for multiple comparison was used in the comparison of cluster analysis results. The letters "a" to "f" indicate a significant difference between two group respectively (a = cluster 1 vs. cluster 2, b = cluster 1 vs. cluster 3, c = cluster 1 vs. cluster 4, d = cluster 2 vs. cluster 3, e = cluster 2 vs. cluster 4, f = cluster 3 vs. cluster 4. PAM, partitioning around medoids; FiO_2, fraction of inspired oxygen; SBP, Systolic blood pressure; DBP, Diastolic blood pressure; ARDS, acute respiratory distress syndrome; ALT, Alanine aminotransferase; AST, Aspartate aminotransferase; CRNMB, clinically relevant non-major bleeding; FOBT, fecal occult blood test; GOBT, gastric occult blood test.

significant coagulopathy, characterized by thrombocytopenia, mildly prolonged prothrombin time, and elevated serum D-dimer levels (27). Recent research indicates that coagulopathy is not only common in COVID-19 patients but is also associated with increased mortality (9). The potential mechanism for the development of coagulopathy in COVID-19 patients may

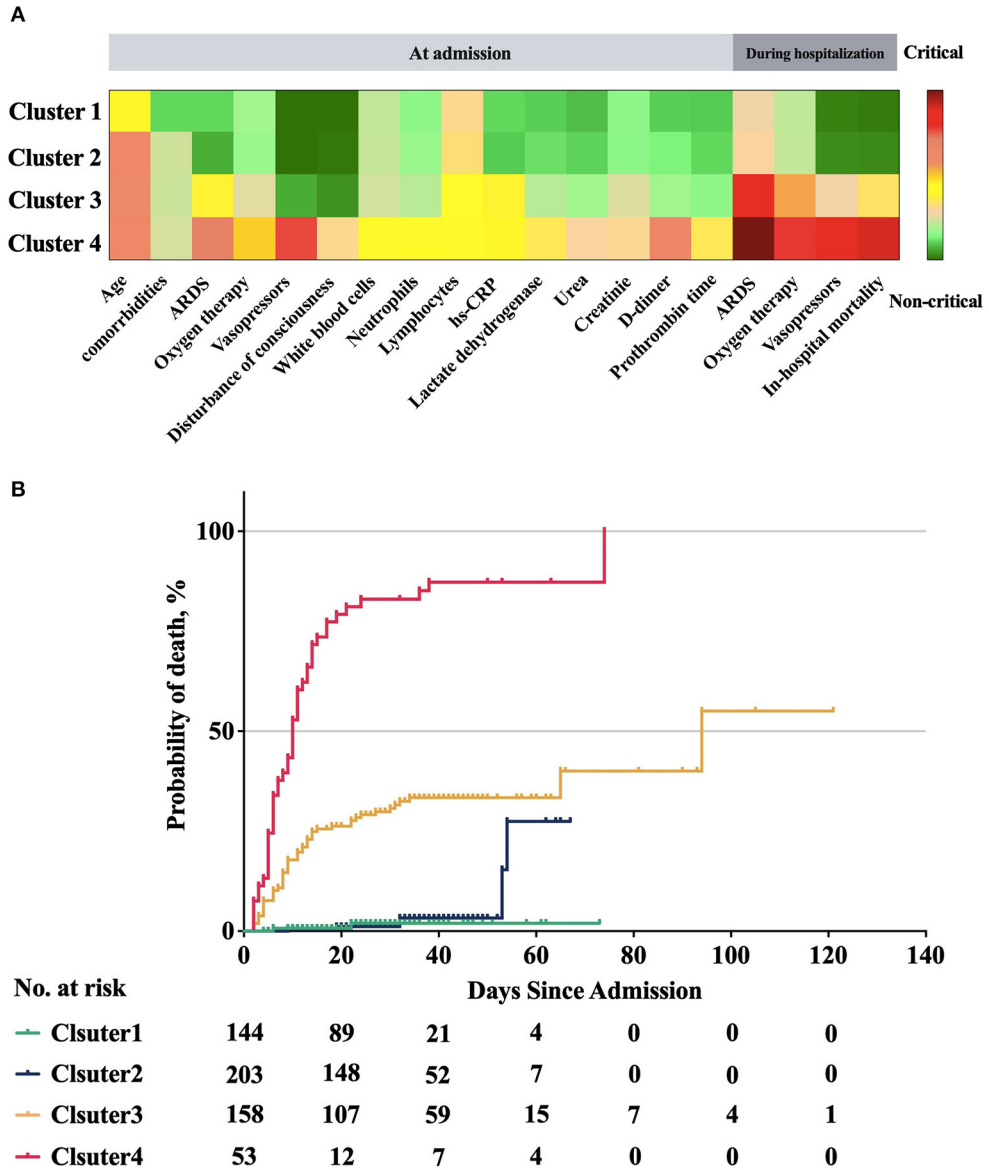


FIGURE 5 | Clinical characteristics and probability of death of the four clusters. Clusters 1 and 2 were non-critical cases from admission to discharge and they showed lower probability of death. Cluster 3 had significant higher in hospital mortality and probability of death than clusters 1 and 2, but lower than that in cluster 4. Cluster 4 had the highest in hospital mortality and probability of death, and there were significant differences with the other three clusters. **(A)** Clinical characteristics of four clusters. Some of the significant clinical features are plotted for each cluster. The features are displayed by color-coded heatmap with normalized values. **(B)** Comparison of probability of death between four clusters. There was statistically significant difference in the survival distribution between any two groups ($P < 0.001$), except that there was no statistically significant difference in the survival distribution of cluster 1 and cluster 2 (Log-Rank test, $P = 0.512 > 0.008$).

be related to endothelial cell dysfunction (37) and hypoxia-induced thrombosis (38) following a SARS-CoV-2 infection. Because the endothelium plays an important role in regulating hemostasis, fibrinolysis, and vessel wall permeability, endothelial dysfunction in pulmonary microvessels may act as a trigger for immunothrombosis, resulting in coagulopathy. Histological analysis of pulmonary vessels in COVID-19 patients shows more widespread thrombosis with microangiopathy compared to that observed in patients with influenza. Based on this preliminary evidence, AC treatment may be beneficial for COVID-19 patients by inhibiting thrombin generation and thereby reducing mortality. The ISTH suggests that a prophylactic dose of LMWH should be considered in all patients without contraindications. Moreover, the Chinese Diagnosis and Treatment Protocol for COVID-19 Patients (Version 8.0) also suggested using AC treatment in selected patients. However, these recommendations require additional clinical evidence to determine the association between AC treatment and the outcome of COVID-19 patients,

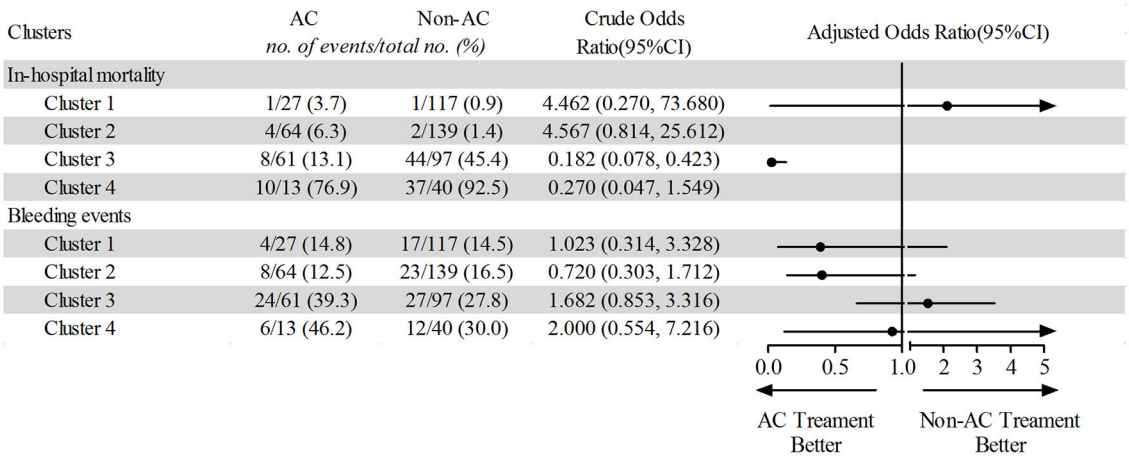


FIGURE 6 | Comparing of in-hospital mortality and bleeding events between AC treatment and non-AC treatment patients based on unsupervised machine learning. AC, anticoagulation.

and also to clarify the indications, contradictions and optimal duration, dose, and time to use AC. We conducted this matched cohort study using a comprehensive source of COVID-19 patients. In general, results showed that receiving AC treatment was associated with a decreased in-hospital mortality in these patients. Although patients who received AC treatment exhibited a significant increase in CRNMB and microscopic hematuria, they had no increase in the incidence of major bleeding. A clinical subgroup analysis was also carried out to identify patient subgroups who receive greater benefit from AC treatment. At hospital admission, patients of severe COVID-19 clinical cases, patients with mild ARDS or patients who had a D-dimer level $\geq$0.5 μg/mL were more likely to benefit from AC therapy. During hospitalization, patients who developed severe ARDS or critical COVID-19 cases were more likely to benefit from AC therapy (**Figure 3**). Results of clusters identified by unsupervised machine learning revealed similar results as of clinical subgroups. Critical patients of cluster 3 could benefit from AC treatment whereas non-critical patients in clusters 1 and 2 did not. Nevertheless, a sub-phenotype (cluster 4) exhibited even severe multiple organ dysfunction and excessive inflammation might not benefit from AC therapy.

To date, several research works have investigated systemic AC therapy in COVID-19 patients (9–13, 39). Although the results generally suggested that AC treatment was associated with lower mortality of COVID-19 patients, a constant instruction for clinic application was not easy to conclude. Several possible reasons are worth to be noted. As a retrospective cohort study, imbalance of baseline covariates, covariates related to outcome and covariates related to exposure assignment might lead to biased results. Among the existing studies, some studies used propensity score methods for reducing the effects of confounding (12, 39), some did not (9, 10). In our study, we applied PSM which yield a relatively balanced cohort. We also did IPTW analysis, another propensity score method, to detect the selective bias potentially caused by PSM in the full cohort. The results from PS matched cohort and IPTW analysis in the full cohort both revealed that AC treatment was associated with lower death risk. We also noted that, without the PS method, either the crude results or baseline-adjusted results will lead to an adverse conclusion. Immortal time is a gap period between exposure (usually the span after cohort follow-up) and initiation of follow-up (20). This might cause potential immortal time bias and exaggerate the association between the exposure and outcome. As a result, we carried out a Cox proportional hazards model with a time-dependent manner for the drug exposure in this study.

Another key question concerns the confounders involving various durations, dosages and types of AC treatment. In a retrospective cohort study from the Mount Sinai Health System (11), the duration of hospitalization (median 5 days, IQR 3–8 days) and the course of AC treatment (median 3 days, IQR 2–7 days) were relatively short. Within the current consensus on anticoagulant therapy for venous thromboembolism, it is generally considered that patients with confirmed deep vein thrombosis or pulmonary embolism need LMWH treatment for at least 5 days followed by Dabigatran or Edoxaban (40). As a new disease without comprehensive study until now, to determine the optimal duration, type and dosage of AC treatment need more evidence. In our study, we conducted a series of sensitivity analyses to investigate the relationship between outcome and AC treatment duration, dosage and type. We found that AC treatment for 7 days or longer was associated with a lower death risk while AC treatment for <7 days was not. Low dose thromboprophylaxis, intermediate dose thromboprophylaxis and therapeutic dose anticoagulation were all associated with lower death risk. Although we recorded detailed AC treatment type, the majority was LMWH (**Supplementary Figure 1**), we only investigated LMWH and non-LMWH for sensitivity analyses here. It is revealed that both LMWH and non-LMWH treatment were associated with a lower death rate.

The heterogeneity of the research population is another vital confounder that influences the results. In our study, we

analyzed a full cohort of unselected patients from two designated hospitals including mild to critical cases. In general, we found that AC treatment was associated with low mortality, which had a constant result with the previous studies. Furtherly, we investigated who might benefit from AC treatment in subgroup analyses. The stratification criteria used in our study included the most frequently used classification of clinical severity of the COVID-19 patients. It is well-known that hypoxia is a core clinical manifestation and major pathophysiology change that contributes to the death of COVID-19 patients (41–43). The classification of both ARDS (18, 29, 44) and COVID-19 clinical severity classification (17, 45) indicates the severity of hypoxia and accordingly, they are used frequently by clinicians to evaluate and triage patients and to decide major treatments (e.g., levels of oxygen therapy). As a result, we stratified patients according to ARDS classification, COVID-19 clinical classification, and D-dimer levels at both hospital admission and during hospitalization. By this strategy, we found that, at admission, severe cases of COVID-19 clinical classification, mild ARDS cases and patients with a D-dimer level $\geq 0.5\,\mu g/mL$ may benefit from AC. While during the hospital stay, critical cases and severe ARDS cases may benefit from AC. These results were in constant with Sun et al.' study (9) with severe cases and subgroup analysis from Shen et al. (39) and CORIST Studies (12).

Clustering the study population may help minimize the influence of heterogeneity on the results. The traditional way to categorize patients is based on pre-defined standards. The standards are usually defined by a group of experienced experts with a strong background and prior knowledge in the medical area. Therefore, the procedure for generating the standards alone takes considerable effort and time. In addition, these standards cannot easily be quickly established or updated for a new situation in a short period, which was apparent when we faced this new pandemic, COVID-19. Unsupervised clustering algorithms in machine learning offer another perspective to perform the identification of data subclasses. Unsupervised clustering approaches can achieve more stable and robust clustering results without any prior knowledge of the meaning of each variable in the data. In addition, it may also identify some intrinsic correlations between the variables which sometimes cannot be easily noticed by human experts. Considering the heterogeneity of COVID-19 patients, an innovative strategy was carried out to identify subphenotype of patients who exhibited distinct clinical characteristics and respond to certain treatment using unsupervised learning approach (46). To this end, a four-class PAM-based clustering model was established, representing four distinct COVID-19 patient subphenotypes with different clinical characteristics. In particular, clusters 1 and 2 were non-critical cases with significantly lower mortality. Patients in these two clusters did not benefit from AC treatment. Clusters 3 and 4 were critical cases both exhibited significant abnormal laboratory testing results at admission and unstable vital sign. Cluster 3 had mild to moderate ARDS at admission and progressed to severe ARDS during the hospital stay. Patients in cluster 3 can benefit from AC treatment and had no significant increase in bleeding events. Compared to cluster 3, cluster 4 was the most critical cases and has the highest mortality. A novel result by the clustering approach was that, among these most critical patients (clusters 4), who had moderate or severe ARDS at admission and developed severe ARDS during the hospital stay, AC treatment was not associated with a lower death risk. Further characteristic analysis of these clusters revealed cluster 4 was characterized by multiple organ dysfunction and excessive inflammation. This led us to conclude that critical COVID-19 patients with these features cannot benefit from AC treatment. Recently, an open-label, adaptive, multiplatform, randomized control trial was published (14), with the researchers noting that the initial strategy of therapeutic-dose anticoagulation did not result in a greater probability of survival in critically ill COVID-19 patients (defined as COVID-19 that led to the receipt of ICU-level respiratory or cardiovascular organ support in an ICU) compared to usual-care thromboprophylaxis. This result was different from our clinical subgroup analysis but similar to the phenotypes of clusters 4 in our unsupervised clustering analysis.

Analysis of safety endpoints showed that although the risk of bleeding events, including CRNMB and microscopic hematuria, were higher in the AC group compared to the non-AC group, there was no significant difference in the risk of major bleeding events or thrombocytopenia between the two groups. In brief, the above findings suggested that the use of AC treatment for 7 days or longer in hospitalized COVID-19 patients was associated with increased CRNMB and microscopic hematuria but not with other bleeding events, especially major bleeding. These key observations are consistent with those reported in recent studies (11, 39). Although the competing risk model analysis in this study revealed that there was no significant difference in bleeding risk between the AC and non-AC groups when considering death as a competitive event, the increase of CRNMB still reminds clinicians should be more cautious when using anticoagulation treatment.

This study had several limitations. Firstly, as a retrospective cohort study, imbalanced confounders and selective bias may exist. Large-scale, multicenter, randomized, controlled trials are urgently needed to fully assess the efficacy of AC in patients with COVID-19. Besides, our unsupervised clustering model did not take the importance of each variable into consideration, as it treated all the variables equally as numerical values and measured the similarity between patients based on geometric distance. However, the variables could have completely different semantic meanings. Therefore, it is still necessary for human experts to inspect the clustering results to make sure that the results are explainable. Future work could be to intrinsically integrate the importance of clinic variables into the similarity measurements of unsupervised cluster models.

CONCLUSION

COVID-19 patients who received AC treatment for 7 days or longer had a significantly lower in-hospital death risk but not higher risk of major bleeding. Through the clinical subgroup analysis, critically ill patients were more likely to benefit from AC treatment. Specifically, the unsupervised machine learning model revealed that, within critically ill COVID-19

patients, patients characterized by multiple organ dysfunction (neurologic, circulation, coagulation, kidney and liver dysfunction) and excessive inflammation may not benefit from AC treatment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

This retrospective cohort study was approved by the ethics committee of Tongji Medical College, Huazhong University of Science and Technology (No. 2020-S220). The clinical trial was registered and verified by the Chinese Clinical Trial Registry (ChiCTR2000039855).

AUTHOR CONTRIBUTIONS

YB, YL, SW, and SL conceptualized the paper. YB, ZH, and SL had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. HD and JC performed the unsupervised clustering analysis by machine learning and established the clustering model. YB, YL, ZH, YFa, GY, SY, YW, JL, YFe, LX, YZ, and ZY contributed to the acquisition, analysis, and interpretation of data. YB and YL drafted the manuscript. HD and JC drafted the clustering model part of the manuscript. PZ and SL helped revise the manuscript. YB, SW, and JH did the statistical analysis. SL and YB obtained the fundings. SL attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. YB and SL supervised the study and are the guarantors. All authors contributed to the critical revision of the manuscript for important intellectual content and gave final approval of the version to be published.

FUNDING

This study was supported by the COVID-19 Rapid Response Research Project of Huazhong University of Science and Technology (Grant 2020kfyXGYJ049, to SL), and the Natural Science Foundation of Hubei Province (Grant No. 2021CFB376, to YB).

ACKNOWLEDGMENTS

The authors thank all the included patients and their families, physicians, nurses, dieticians, and all staff.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.786414/full#supplementary-material

REFERENCES

1. Asselah T, Durantel D, Pasmant E, Lau G, Schinazi RF. COVID-19: discovery, diagnostics and drug development. *J Hepatol.* (2021) 74:168–84. doi: 10.1016/j.jhep.2020.09.031
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.* (2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5
3. Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, et al. Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *N Engl J Med.* (2020) 383:120–8. doi: 10.1056/NEJMoa2015432
4. Fox SE, Akmatbekov A, Harbert JL Li G, Quincy Brown J, Vander Heide RS. Pulmonary and cardiac pathology in African American patients with COVID-19: an autopsy series from New Orleans. *Lancet Respir Med.* (2020) 8:681–6. doi: 10.1016/S2213-2600(20)30243-5
5. Magro C, Mulvey JJ, Berlin D, Nuovo G, Salvatore S, Harp J, et al. Complement associated microvascular injury and thrombosis in the pathogenesis of severe COVID-19 infection: a report of five cases. *Transl Res.* (2020) 220:1–13. doi: 10.1016/j.trsl.2020.04.007
6. Ding Y, Wang H, Shen H, Li Z, Geng J, Han H, et al. The clinical pathology of severe acute respiratory syndrome (SARS): a report from China. *J Pathol.* (2003) 200:282–9. doi: 10.1002/path.1440
7. Nicholls JM, Poon LLM, Lee KC, Ng WF, Lai ST, Leung CY, et al. Lung pathology of fatal severe acute respiratory syndrome. *Lancet.* (2003) 361:1773–8. doi: 10.1016/S0140-6736(03)13413-7
8. Ng DL, Al Hosani F, Keating MK, Gerber SI, Jones TL, Metcalfe MG, et al. Clinicopathologic, immunohistochemical, and ultrastructural findings of a fatal case of middle east respiratory syndrome coronavirus infection in the United Arab Emirates, April 2014. *Am J Pathol.* (2016) 186:652–8. doi: 10.1016/j.ajpath.2015.10.024
9. Tang N, Bai H, Chen X, Gong J, Li D, Sun Z. Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *J Thromb Haemost.* (2020) 18:1094–9. doi: 10.1111/jth.14817
10. Ayerbe L, Risco C, Ayis S. The association between treatment with heparin and survival in patients with Covid-19. *J Thromb Thrombolysis.* (2020) 50:298–301. doi: 10.1007/s11239-020-02162-z
11. Paranjpe I, Fuster V, Lala A, Russak AJ, Glicksberg BS, Levin MA, et al. Association of treatment dose anticoagulation with in-hospital survival among hospitalized patients with COVID-19. *J Am Coll Cardiol.* (2020) 76:122–4. doi: 10.1016/j.jacc.2020.05.001
12. Di Castelnuovo AF, Costanzo S, Iacoviello L. Heparin in COVID-19 patients is associated with reduced in-hospital mortality: the multicentre Italian CORIST Study. *Thromb Haemost.* (2021) 121:1054–65. doi: 10.1055/a-1347-6070
13. Evans EL, Povstyan OV, De Vecchis D, Macrae F, Lichtenstein L, Futers TS, et al. Impact of anticoagulation prior to COVID-19 infection: a propensity score–matched cohort study. *Blood.* (2020) 136:140–4. doi: 10.1182/blood.2019004174
14. Investigators R-C. Investigators ACA, Investigators A, Goligher EC, Bradbury CA, McVerry BJ, et al. Therapeutic Anticoagulation with Heparin in Critically Ill Patients with Covid-19. *N Engl J Med.* (2021) 385:777–89. doi: 10.1056/NEJMoa2103417
15. American Society of Hematology. *COVID-19 and VTE/Anticoagulation: Frequently Asked Questions*. (2021). Available online at: https://www.hematology.org/%20covid-19/covid-19-and-vte-anticoagulation (accessed April 12, 2021).
16. National Institutes of Health. *COVID-19 Treatment Guidelines*. (2021). Available online at: https://www.covid19treatmentguidelines.nih.gov./antithrombotic-therapy/ (accessed April 12, 2021).

17. National Health Commission of the People's Republic of China. *Diagnosis and Treatment Protocol for COVID-19 Patients (Trial Version 8)*. (2020). Available online at: http://www.nhc.gov.cn/yzygj/s7653p/202008/0a7bdf12bd4b46e5bd28ca7f9a7f5e5a.shtml (accessed August 19, 2020).
18. Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, et al. Acute respiratory distress syndrome: the Berlin Definition. *JAMA*. (2012) 307:2526–33. doi: 10.1001/jama.2012.5669
19. World Health Organization. *Clinical Management of Severe Acute Respiratory Infection When Novel Coronavirus (2019-nCoV) Infection Is Suspected: Interim Guidance*. (2020). Available online at: https://www.who.int/publications/i/item/10665-332299 (accessed April 12, 2021).
20. Suissa S. Immortal time bias in pharmaco-epidemiology. *Am J Epidemiol.* (2008) 167:492–9. doi: 10.1093/aje/kwm324
21. Schulman S, Kearon C; Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost*. (2004) 3:692–4. doi: 10.1111/j.1538-7836.2005.01204.x
22. Kaatz S, Ahmad D, Spyropoulos AC, Schulman S. Subcommittee on Control of Anticoagulation. Definition of clinically relevant non-major bleeding in studies of anticoagulants in atrial fibrillation and venous thromboembolic disease in non-surgical patients: communication from the SSC of the ISTH. *J Thromb Haemost.* (2015) 13:2119–26. doi: 10.1111/jth.13140
23. Thachil J, Tang N, Gando S, Falanga A, Cattaneo M, Levi M, et al. ISTH interim guidance on recognition and management of coagulopathy in COVID-19. *J Thromb Haemost.* (2020) 18:1023–6. doi: 10.1111/jth.14810
24. Neill AM, Martin IR, Weir R, Anderson R, Chereshsky A, Epton MJ, et al. Community acquired pneumonia: aetiologyand usefulness ofseveritycriteriaon admission. *Thorax*. (1996) 51:1010–6. doi: 10.1136/thx.51.10.1010
25. Ware LB, Matthay MA. The acute respiratory distress syndrome. *New Engl J Med.* (2000) 342:1334–49. doi: 10.1056/NEJM200005043421806
26. Riviello ED, Kiviri W, Twagirumugabe T, Mueller A, Banner-Goodspeed VM, Officer L, et al. Hospital incidence and outcomes of ARDS using the Kigali modification of the Berlin definition. *Am J Respir Crit Care Med*. (2016) 193:52–9. doi: 10.1164/rccm.201503-0584OC
27. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The Third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287
28. Aggarwal CC, Hinneburg A, Keim DA. On the surprising behavior of distance metrics in high dimensional space. In: *International Conference on Database Theory*. Berlin: Springer (2001). p. 420–34.
29. Weakliem, David L. A critique of the Bayesian information criterion for model selection. *Sociol Methods Res*. (1999) 27:359–97. doi: 10.1177/0049124199027003002
30. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res.* (2011) 46:399–424. doi: 10.1080/00273171.2011.568786
31. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax*. (2003) 58:377–82. doi: 10.1136/thorax.58.5.377
32. Kuss O, Blettner M, Borgermann J. Propensity score: an alternative method of analyzing treatment effects. *Dtsch Arztebl Int.* (2016) 113:597–603. doi: 10.3238/arztebl.2016.0597
33. Geleris J, Sun Y, Platt J, Zucker J, Baldwin M, Hripcsak G, et al. Observational study of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl J Med.* (2020) 382:2411–8. doi: 10.1056/NEJMoa2012410
34. Chen C, Chen C, Yan JT, Zhou N, Zhao JP, Wang DW. Analysis of myocardial injury in patients with COVID-19 and association between concomitant cardiovascular diseases and severity of COVID-19. *Zhonghua Xin Xue Guan Bing Za Zhi.* (2020) 48:567–71. doi: 10.3760/cma.j.cn112148-20200225-00123
35. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, et al. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med.* (2020) 180:1081–9. doi: 10.1001/jamainternmed.2020.2033
36. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med.* (2020) 180:934–43. doi: 10.1001/jamainternmed.2020.0994
37. Gando S, Levi M, Toh CH. Disseminated intravascular coagulation. *Nat Rev Dis Primers.* (2016) 2:16037. doi: 10.1038/nrdp.2016.37
38. Gupta N, Zhao YY, Evans CE. The stimulation of thrombosis by hypoxia. *Thromb Res.* (2019) 181:77–83. doi: 10.1016/j.thromres.2019.07.013
39. Shen L, Qiu L, Liu D, Wang L, Huang H, Ge H, et al. The association of low molecular weight heparin use and in-hospital mortality among patients hospitalized with COVID-19. *Cardiovasc Drugs Ther.* (2021) 4:1–8. doi: 10.1007/s10557-020-07133-3
40. Schulman S, Konstantinides S, Hu Y, Tang LV. Venous thromboembolic diseases: diagnosis, management and thrombophilia testing: observations on NICE guideline [NG158]. *Thromb Haemost.* (2020) 120:1143–6. doi: 10.1055/s-0040-1712913
41. Somers VK, Kara T, Xie J. Progressive hypoxia: a pivotal pathophysiologic mechanism of COVID-19 pneumonia. *Mayo Clin Proc.* (2020) 95:2339–42. doi: 10.1016/j.mayocp.2020.09.015
42. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline characteristics and outcomes of 1591 patients infected with SARS-CoV-2 admitted to ICUs of the Lombardy Region, Italy. *JAMA.* (2020) 323:1574–81. doi: 10.1001/jama.2020.5394
43. Xie J, Covassin N, Fan Z, Singh P, Gao W, Li G, et al. Association between hypoxemia and mortality in patients with COVID-19. *Mayo Clin Proc.* (2020) 95:1138–47. doi: 10.1016/j.mayocp.2020.04.006
44. Ferrando C, Suarez-Sipmann F, Mellado-Artigas R, Hernandez M, Gea A, Arruti E, et al. Clinical features, ventilatory management, and outcome of ARDS caused by COVID-19 are similar to other causes of ARDS. *Intensive Care Med.* (2020) 46:2200–11. doi: 10.1007/s00134-020-06251-8
45. Nicastri E, Petrosillo N, Ascoli Bartoli T, Lepore L, Mondi A, Palmieri F, et al. National Institute for the Infectious Diseases "L. Spallanzani", IRCCS Recommendations for COVID-19 clinical management. *Infect Dis Rep.* (2020) 12:8543. doi: 10.4081/idr.2020.8543
46. Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, et al. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med.* (2014) 2:611–20. doi: 10.1016/S2213-2600(14)70097-9

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 17 January 2022
doi: 10.3389/fmed.2021.728521

Machine Learning Prediction Model for Acute Renal Failure After Acute Aortic Syndrome Surgery

Jinzhang Li[1,2,3], Ming Gong[2,3], Yashutosh Joshi[4], Lizhong Sun[2], Lianjun Huang[5], Ruixin Fan[6], Tianxiang Gu[7], Zonggang Zhang[8], Chengwei Zou[9], Guowei Zhang[10], Ximing Qian[11], Chenhui Qiao[12], Yu Chen[13], Wenjian Jiang[2,3,4]* and Hongjia Zhang[1,2,3]**

[1] Department of Physiology and Pathophysiology, School of Basic Medical Sciences, Capital Medical University, Beijing, China, [2] Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing, China, [3] Beijing Lab for Cardiovascular Precision Medicine, Beijing, China, [4] Department of Cardiothoracic Surgery, St Vincent's Hospital, Sydney, NSW, Australia, [5] Department of Interference Diagnosis and Treatment, Beijing Anzhen Hospital, Capital Medical University, Beijing, China, [6] Department of Cardiovascular Surgery, Guangdong Provincial People's Hospital, Guangzhou, China, [7] Department of Cardiac Surgery, First Affiliated Hospital, China Medical University, Shenyang, China, [8] Department of Cardiac Surgery, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China, [9] Department of Cardiovascular Surgery, Shandong Provincial Hospital Affiliated With Shandong First Medical University, Jinan, China, [10] Department of Cardiovascular Surgery, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [11] Department of Cardiac Surgery, School of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China, [12] Department of Cardiovascular Surgery, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [13] Department of Cardiac Surgery, Peking University People's Hospital, Beijing, China

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Kevin J. Ni,
St George Hospital, Australia
Yi Yang,
Southeast University, China

***Correspondence:**
Hongjia Zhang
zhanghongjia722@ccmu.edu.cn
Wenjian Jiang
jiangwenjian@ccmu.edu.cn
Ming Gong
gongmaster@126.com

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: 21 June 2021
Accepted: 24 December 2021
Published: 17 January 2022

Citation:
Li J, Gong M, Joshi Y, Sun L, Huang L, Fan R, Gu T, Zhang Z, Zou C, Zhang G, Qian X, Qiao C, Chen Y, Jiang W and Zhang H (2022) Machine Learning Prediction Model for Acute Renal Failure After Acute Aortic Syndrome Surgery. Front. Med. 8:728521. doi: 10.3389/fmed.2021.728521

Background: Acute renal failure (ARF) is the most common major complication following cardiac surgery for acute aortic syndrome (AAS) and worsens the postoperative prognosis. Our aim was to establish a machine learning prediction model for ARF occurrence in AAS patients.

Methods: We included AAS patient data from nine medical centers (n = 1,637) and analyzed the incidence of ARF and the risk factors for postoperative ARF. We used data from six medical centers to compare the performance of four machine learning models and performed internal validation to identify AAS patients who developed postoperative ARF. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was used to compare the performance of the predictive models. We compared the performance of the optimal machine learning prediction model with that of traditional prediction models. Data from three medical centers were used for external validation.

Results: The eXtreme Gradient Boosting (XGBoost) algorithm performed best in the internal validation process (AUC = 0.82), which was better than both the logistic regression (LR) prediction model (AUC = 0.77, $p < 0.001$) and the traditional scoring systems. Upon external validation, the XGBoost prediction model (AUC =0.81) also performed better than both the LR prediction model (AUC = 0.75, $p = 0.03$) and the traditional scoring systems. We created an online application based on the XGBoost prediction model.

Conclusions: We have developed a machine learning model that has better predictive performance than traditional LR prediction models as well as other existing risk scoring systems for postoperative ARF. This model can be utilized to provide early warnings when high-risk patients are found, enabling clinicians to take prompt measures.

Keywords: machine learning, acute renal failure, acute aortic syndrome, prediction model, eXtreme Gradient Boosting

INTRODUCTION

Acute aortic syndrome (AAS) is a serious and life-threatening disease process involving the ascending aorta and aortic arch. Traditionally, surgical intervention is the best way to treat AAS (1). Acute renal failure (ARF) is an important complication affecting the prognosis of AAS patients after surgery. This complication indicates that the patient has a poor prognosis, and it can increase postoperative mortality and morbidity (2). While renal replacement therapy (RRT) is a feasible treatment modality, it is arguably more important to identify the risk factors for postoperative ARF and identify potential patients with a higher likelihood of developing ARF in the postoperative setting. Some scoring systems already exist for predicting ARF after cardiac surgery (3–6), but they are usually employed for coronary artery bypass graft or heart valve surgery. Whether these scoring systems can be used in AAS-related surgery is unclear.

In recent years, machine learning has become increasingly widely used in medicine; it can help us process large amounts of data and find potential data relationships. Multiple excellent algorithms have been developed in the field of machine learning so that we can use them to build predictive models.

The main purpose of this study was to establish a predictive model for the occurrence of ARF in AAS patients after surgery through machine learning, thereby helping to identify potential patients who may develop ARF, and compare it with a traditional logistic regression (LR) prediction model and other scoring systems. This study followed the recommendations of the *Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis* statement (7).

MATERIALS AND METHODS

Participants

A total of 1,637 AAS patients undergoing surgery and treatment at nine medical centers in China from January 1, 2015, to December 31, 2019 were recruited for this study. The ethics committee of Beijing Anzhen Hospital approved this retrospective cohort study (No. 2018015; Date: 2018-10-18). Patients' written informed consent was waived due to the retrospective nature of the study. We collected demographic, surgical, and clinical data with a potential relationship to the renal function of patients from admission through discharge. Patients who had renal failure before surgery, incomplete surgical data, or surgery involving the abdominal aorta and below were excluded. All patients were diagnosed with Stanford type-A AAS through aortic computed tomography angiography (CTA) by experienced imaging specialists and cardiovascular surgeons. The diagnosis of ARF was established according to the Kidney Disease: Improving Global Outcomes guidelines (8). Postoperative ARF was defined as an increase of >3 times or an increase of >4.0 mg/dL (353.6 μmol/L) in postoperative serum creatinine (Scr) or the initiation of RRT compared to baseline. The estimated glomerular filtration rate (eGFR) was calculated using the Chronic Kidney Disease Epidemiology Collaboration formula (CKD-EPI) (9). Surgery was performed by the surgical team of the medical center at the time of the patient's admission.

Surgical Details

Anesthesia was maintained by either total intravenous anesthetics (propofol and sufentanil) or an inhalational agent (sevoflurane) with vecuronium bromide. Tranexamic acid was used for coagulation support. Cardiopulmonary bypass (CPB) was routinely instituted at 2.2 to 2.5 L/min/m^2. When the lesion involved the aortic arch, arterial cannulation was performed in the right axillary and/or femoral artery and/or ascending aorta; venous cannulations were bicaval. Cold blood cardioplegia for myocardial protection was perfused through the left and right coronary arteries. If the distal aorta or aortic arch needed reconstruction, this process was performed under deep or moderate hypothermia and circulatory arrest. Once the distal reconstruction was complete, the aortic graft was clamped proximally. Selective anterograde perfusion was most often instituted through the innominate arteries. During core cooling, accompanying cardiac procedures, including aortic valve repair or replacement, sinus reconstruction, and root replacement, were performed if necessary. If the lesion involved only the ascending aorta, arterial cannulation was performed in the ascending aorta; venous cannulations were bicaval. Subsequently, reconstruction of the ascending aorta was performed.

Data Pre-processing

For missing data, we used the k-nearest neighbors approach to fill in missing values (10). By calculating the Euclidean distance between each case, the missing value was imputed using the mean value from the five nearest neighbors. When the data were in the range of 0 to 1, most machine learning algorithms had excellent performance. To improve the performance of machine learning, we used the method provided by the MinMaxScaler function to scale the data during data pre-processing.

Statistical Analysis

The description of the data and basic statistical analysis were performed using IBM SPSS Statistics for Windows Version 25.0. Continuous variables are expressed as the median (along with

the first and third quartile values). Categorical variables are expressed as frequencies (n) with percentages (%). Statistical analysis of continuous variables was performed using the Mann–Whitney U test, while categorical variables were analyzed using the chi-squared test and Fisher's exact test. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve was compared with the machine learning prediction model. DeLong's test (11) was used to calculate the P value. A $P < 0.05$ was considered statistically significant, and all statistical tests were two-sided.

Establishment of Prediction Model

We used Python for programming and used the scikit-learn 0.22.1 package to build machine learning classifiers (12). The concise process of establishing and evaluating the prediction model is shown in **Supplementary Figure 1**.

The machine learning prediction model used data from 1,637 patients from nine medical centers in China. Among them, 1,318 patient data points from six medical centers in Beijing, Zhejiang, Shandong, Liaoning and Henan provinces were used for machine learning training and internal validation. Training and validation data were divided by ten-fold cross-validation, each time 90% of the data was used as training data, and 10% of the data was used as validation data. And 319 patient data points from three medical centers in Heilongjiang and Guangdong provinces and Xinjiang Uygur Autonomous Region were used for external validation of the prediction model. The division of internal validation and external validation data was determined by the geographic location of the medical centers. The six medical centers used for machine learning training and internal validation were located in central China, and the three medical centers used for external validation were located in northern, southern and western China. This division method was suitable for evaluating the generalization ability of predictive models.

For the selection of machine learning algorithms, we chose the support vector machine classifier (SVC) linear kernel and Nu-SVC with the radial basis function kernel in the SVC algorithm. These are two classic algorithms that use the classifier with the largest interval in the feature space for classification, and can perform data classification after linear-range or high-dimensional mapping. It is still valid when the feature has a high-dimensional relationship. Among them, SVC uses a linear algorithm, while Nu-SVC uses a radial basis function. At the same time, we chose the AdaBoost algorithm (13) and the eXtreme Gradient Boosting (XGBoost) algorithm (14) in the ensemble methods, which are popular algorithms in machine learning classifiers and can combine the predictions of several base estimators so that they have excellent performance. AdaBoost integrates multiple basic decision trees, uses misclassified data points to identify problems, and improves the model by adjusting the weights of misclassified data points. XGBoost uses negative gradients to identify problems, and calculates negative gradients to improve the model. In addition, we also tested the combination of two algorithms. This combination uses two independent algorithms to build prediction models separately, and uses the results of the two prediction models as features to retrain the new prediction model, which is also called a stacking algorithm. We tested the combination of XGBoost + random forest algorithm and XGBoost + decision tree algorithm.

Feature Selection

To make the prediction model more accurate, we selected all demographic characteristics and preoperative clinical data as the features for machine learning. At the beginning of model training, all 134 features were used (**Supplementary Table 3**), and Shapley additive explanations (SHAP) was used to judge the importance of each feature. In the machine learning prediction model, SHAP can analyse the impact of each feature of each patient on the prediction result (15). Finally, the features that were considered important in all prediction models were used as the final features of the machine learning model.

Ten-Fold Cross-Validation

Ten-fold cross-validation is considered a reliable method for model evaluation and performance improvement (16), and it was used for parameter adjustment and algorithm comparison. Since machine learning algorithms usually cannot use training data as test data, 10-fold cross-validation is generally used to evaluate machine learning algorithms. Ten-fold cross-validation can divide the data into 10 parts. The classifier used nine of them for training, and the remaining part was used for testing. Ten repetitions constituted a 10-fold cross-validation (**Supplementary Figure 2**). The average of 10 test results was used to evaluate the predictive ability of machine learning algorithms and parameters.

Each classifier algorithm had many parameter settings, and the choice of parameters had a great impact on the results of the classifier. We used the grid-search algorithm and internal validation data to determine the best parameters for each classifier algorithm.

The grid-search algorithm used 10-fold cross-validation to select the parameters of the machine learning algorithm. We told the grid search algorithm the potential optimal parameter range of the classifier, and the grid search algorithm used 10-fold cross validation to calculate the predictive ability of each set of parameters. After the grid-search algorithm calculated each set of parameter combinations, it told us the optimal parameter combination. After constantly changing the parameter range used by the grid search algorithm, the optimal parameter combination of this machine learning algorithm was finally obtained.

Similarly, 10-fold cross-validation was used to compare classifier algorithms. The ROC curve and AUC were calculated at each validation, and the mean and standard deviation of each AUC were compared to obtain the optimal classifier algorithm.

Evaluation of Predictive Models

To compare the machine learning prediction model with the traditional prediction model, we used multivariable LR analysis to establish an LR prediction model with ARF as the end point. In addition, the Cleveland scoring system (3), the simplified renal index (SRI) scoring system (5) and the Leicester scoring system (6) were selected as representatives of the traditional prediction model. The endpoints of these three classic scoring

systems were renal failure and RRT, and they were also scored for complex surgery. To evaluate the prediction model, we compared the machine learning prediction model with the four traditional prediction models. The ROC curve and AUC of the prediction model were calculated using internal validation data. The machine learning prediction model used 10-fold cross-validation to calculate the mean ROC curve and the AUC. In the external validation, the machine learning predictive model was trained using internal validation data. Then, we used the parametric approach based on Platt's logistic model to calibrate the probability of the machine learning model (17) and evaluated the discrimination and calibration of the model by calculating the Brier score. The trained machine learning prediction model was compared with the traditional prediction models using external validation data to evaluate the generalization ability of the prediction model.

RESULTS

Patient Characteristics

A total of 1,637 patients were enrolled in the study, with 1,318 of these cases being used for machine learning training and internal validation. The main characteristics of patients in the internal validation group are presented in **Table 1**. The median age of the patients was 50.0 (42.0–57.0) years; 301 (22.8%) patients were female.

Incidence and Prognosis of Postoperative ARF

The incidence of ARF after aortic surgery was 11.5% (151 in 1,318). The prognostic characteristics of the patients are presented in **Supplementary Table 1**. Patients with postoperative ARF had a poor prognosis and had longer ICU stays (204.0 (104.5–308.2) h vs. 43.0 (20.0–112.5) h, $P < 0.001$) as well as longer ventilator use times (114.0 (62.0–179.0) h vs. 20.0 (15.0–48.0) h, $P < 0.001$). Postoperative ARF may be related to the use of more blood products and drug infusions. Furthermore, patients with postoperative ARF had more postoperative complications (74.8 vs. 34.5%, $P < 0.001$). Most importantly, there were significant differences in mortality between patients with and without ARF (12.6 vs. 0.8%, respectively, $P < 0.001$).

Risk Factors for Postoperative ARF

As a comparison with machine learning models, we used traditional statistical methods to analyse all preoperative and intraoperative factors that either had significant differences or were clinically believed to be related to ARF and calculated the risk factors for postoperative ARF. A multivariable binary LR with the "Forward: LR" method was conducted to determine the risk factors for postoperative ARF (**Table 2**). The results showed that, among the preoperative factors, older age, a higher pulse rate, emergency surgery, and an increased absolute value of leukocytes in the preoperative setting were all risk factors. It was also noted that an increased estimated glomerular filtration rate (eGFR) and platelet count were protective factors against postoperative ARF. In the combined analysis of preoperative and intraoperative factors, in addition to the aforementioned preoperative factors, longer cardiopulmonary bypass time, lower rectal temperature when circulatory arrest, and surgery with circulatory arrest were risk factors for postoperative ARF. We used preoperative factors to establish an LR prediction model for the predictive model to have the ability to predict postoperative ARF of patients before surgery (**Table 2**).

Machine Learning Prediction Model

Feature Selection

In the initial stage, we built machine learning models using all the preoperative features and unoptimized parameters. We analyzed the feature importance of these machine learning models through SHAP (15) and finally selected 15 features for building machine learning prediction models (**Table 3**).

Among the 15 features used to establish these models, we collected complete demographic and renal function data. However, as AAS patients may require emergency surgery, occasionally, the blood test results were partially missing. We used the k-nearest neighbors approach to fill in missing values. **Supplementary Table 2** shows the details of missing values.

Internal Validation

Machine learning models were trained using internal validation data, and the performance of the machine learning models was evaluated using 10-fold cross-validation. The results showed the mean ROC curve and AUC of each machine learning model after 10-fold cross-validation (**Figure 1**). We found that among the prediction models established by a single algorithm, the XGBoost machine learning model performed best (AUC = 0.82, 95% confidence interval (CI): 0.79–0.85), and the combination of XGBoost and other algorithms did not improve performance (**Supplementary Figure 3**); thus, we chose the XGBoost model as the final machine learning model to evaluate its performance. This model had 750 gradient boosted trees, the maximum tree depth was eight, the learning rate was 0.01, the subsample ratio of columns when constructing each tree was 0.75, and the subsample ratio of the training instance was 0.68.

Subsequently, the importance of each feature of the XGBoost model was analyzed by the SHAP method. **Figure 2** shows the results of the feature importance analysis, with more important features distributed on the top and relatively unimportant features on the bottom. Most of the characteristics, either positively or negatively, correlated with the prediction results; however, activated partial thromboplastin time (APTT), fasting blood glucose, body mass index (BMI), international normalized ratio (INR), and alanine aminotransferase (ALT) that were either too high or too low increased the risk of ARF. At the same time, we also analyzed the feature importance based on the fitted trees of the XGBoost model (**Supplementary Figure 4**), and the result is similar to the result of SHAP.

We used internal validation data to calculate the ROC curve and the AUC of the four traditional prediction models (**Figure 3**) and found that the XGBoost model (AUC = 0.82, 95% CI: 0.79–0.85) performed better than the LR prediction model (AUC = 0.77, 95% CI: 0.73–0.81, P<0.001), the Cleveland scoring system (AUC = 0.73, 95% CI: 0.69–0.77, P<0.001), the SRI scoring system (AUC = 0.72, 95% CI: 0.68–0.76, P<0.001), and the

TABLE 1 | Main characteristics of patients in the internal validation groups.

	Overall	Without ARF	Combined ARF	*P* value
Number of patients (cases)	1,318	1,167	151	
Gender, female (cases)	301 (22.8%)	269 (23.1%)	32 (21.2%)	0.61
Age (years)	50.0 (42.0–57.0)	49.0 (41.0–57.0)	53.0 (46.0–60.0)	< 0.001
Information on admission				
Pulse (beats/min)	80.0 (75.0–85.0)	80.0 (75.0–85.0)	80.0 (77.0–88.0)	0.003
Height (cm)	170.0 (166.0–175.0)	170.0 (166.0–175.0)	170.0 (165.0–175.0)	0.06
Weight (kg)	75.0 (65.0–82.4)	75.0 (65.0–83.0)	73.2 (65.0–80.0)	0.62
Body mass index (kg/m^2)	25.4 (22.9–27.8)	25.4 (22.9–27.8)	25.4 (23.4–28.0)	0.35
Systolic pressure (mmHg)	130.0 (120.0–142.0)	130.0 (120.0–143.0)	129.0 (110.0–140.0)	0.009
Diastolic pressure (mmHg)	78.0 (70.0–84.1)	78.0 (70.0–85.0)	78.0 (70.0–82.0)	0.10
Medical history				
Smoking history (cases)	507 (38.5%)	449 (38.5%)	58 (38.4%)	0.99
History of previous cardiac surgery (cases)	107 (8.1%)	98 (8.4%)	9 (6.0%)	0.30
Peripheral vascular disease history (cases)	9 (0.7%)	9 (0.8%)	0 (0.0%)	0.61
Echocardiographic results				
Left ventricular ejection fraction (%)	63.0 (60.0–66.0)	63.0 (60.0–66.0)	62.0 (58.0–65.0)	0.21
Preoperative laboratory examination results				
Absolute value of leukocytes (10^9/L)	9.69 (7.00–13.18)	9.30 (6.80–12.90)	12.33 (10.29–15.33)	< 0.001
Platelets (10^9/L)	185.0 (147.0–228.3)	189.0 (150.0–231.0)	158.0 (126.0–201.0)	< 0.001
Hemoglobin (g/L)	137.0 (122.0–148.0)	137.0 (123.0–148.0)	136.0 (121.0–144.0)	0.12
CK–MB (ng/mL)	1.88 (0.90–9.30)	1.70 (0.80–8.20)	7.10 (1.60–15.00)	< 0.001
Lactate dehydrogenase (U/L)	221.5 (179.0–288.3)	214.0 (176.0–279.0)	277.0 (225.0–332.2)	< 0.001
D–dimer (ng/mL)	1,100.0 (270.8–3,269.3)	940.0 (231.0–2,887.0)	3,328.0 (1,120.0–14,485.0)	< 0.001
INR	31.9 (28.9–36.5)	31.9 (28.8–36.3)	32.2 (29.5–37.8)	0.02
APTT (s)	48.8 (39.6–60.1)	48.6 (39.5–60.0)	51.4 (40.5–65.1)	0.03
Blood amylase (U/dL)	21.0 (15.0–34.0)	21.0 (15.0–33.0)	27.0 (16.0–45.0)	0.07
ALT (U/mL)	22.0 (18.0–32.0)	22.0 (17.0–30.0)	29.0 (21.0–46.0)	0.005
AST (U/mL)	39.1 (35.6–42.1)	39.2 (35.6–42.2)	38.9 (35.2–40.7)	< 0.001
Albumin (g/mL)	78.3 (64.6–99.6)	76.7 (63.9–95.8)	99.7 (78.0–138.6)	0.08
Creatinine (μmol/L)	6.30 (4.99–8.10)	6.10 (4.90–7.78)	8.20 (6.01–10.30)	< 0.001
BUN (mmol/mL)	95.0 (73.1–107.1)	97.1 (77.0–108.4)	69.1 (49.0–93.4)	< 0.001
eGFR (ml/min/1.73 m^2)	6.49 (5.39–7.77)	6.38 (5.30–7.67)	7.26 (6.28–8.48)	< 0.001
Fasting blood glucose (mmol/L)	9.69 (7.00–13.18)	9.30 (6.80–12.90)	12.33 (10.29–15.33)	< 0.001
Diagnosis				
Coronary artery disease (cases)	35 (2.7%)	33 (2.8%)	2 (1.3%)	0.42
Congestive heart failure (cases)	26 (2.0%)	23 (2.0%)	3 (2.0%)	1.00
Chronic respiratory disease (cases)	31 (2.4%)	28 (2.4%)	3 (2.0%)	1.00
Hypertension (cases)	905 (68.7%)	786 (67.4%)	119 (78.8%)	0.004
Diabetes (cases)	62 (4.7%)	54 (4.6%)	8 (5.3%)	0.71
Surgery				
Operative duration (min)	405.0 (340.0–479.0)	396.0 (330.0–465.0)	454.6 (390.0–520.0)	< 0.001
Emergency surgery (cases)	650 (49.3%)	536 (45.9%)	114 (75.5%)	< 0.001
Cardiopulmonary bypass time (min)	186.0 (144.0–224.0)	181.0 (140.0–218.0)	224.0 (188.0–266.0)	< 0.001
Aortic cross–clamp time (min)	104.0 (82.6–131.0)	102.0 (80.0–127.0)	125.0 (103.0–149.0)	< 0.001
With circulatory arrest (cases)	997 (75.6%)	855 (73.3%)	142 (94.0%)	< 0.001
Circulatory arrest time (min)	21.0 (17.4–27.0)	21.0 (17.6–26.6)	21.0 (17.0–27.0)	0.88
Nasopharyngeal temperature when circulatory arrest (°C)	24.1 (23.1–24.9)	24.1 (23.2–24.9)	23.5 (22.5–24.5)	< 0.001
Rectal temperature when circulatory arrest (°C)	25.6 (24.8–26.7)	25.8 (24.9–26.8)	25.0 (24.0–26.0)	< 0.001
RBC transfusion volume (U)	4.00 (0.00–6.00)	3.50 (0.00–6.00)	5.50 (4.00–8.00)	< 0.001

INR, international normalized ratio; APTT, activated partial thromboplastin time; ALT, alanine aminotransferase; AST, aspartate transaminase; BUN, blood urea nitrogen; eGFR, estimated glomerular filtration rate.

TABLE 2 | Multivariable binary logistic regression results.

Characteristics	B	Standard error	*P* value	OR value	OR 95%CI
Preoperative factors only					
Age (years)	0.028	0.010	0.003	1.029	1.010–1.048
Absolute value of leukocyte	0.076	0.022	0.001	1.079	1.033–1.127
Pulse (beats/min)	0.021	0.008	0.006	1.021	1.006–1.037
eGFR (ml/min/1.73m^2)	−0.019	0.004	< 0.001	0.982	0.974–0.989
Platelet (10^9/L)	−0.005	0.002	0.003	0.995	0.992–0.998
Emergency surgery	0.779	0.216	< 0.001	2.179	1.426–3.331
Constant	−4.207	1.057	< 0.001	0.015	
Preoperative and intraoperative factors					
Age (years)	0.032	0.010	0.001	1.032	1.012–1.052
Absolute value of leukocyte	0.062	0.023	0.008	1.064	1.017–1.114
Pulse (beats/min)	0.023	0.008	0.004	1.023	1.007–1.039
eGFR (ml/min/1.73m^2)	−0.017	0.004	< 0.001	0.983	0.976–0.991
Platelet (10^9/L)	−0.004	0.002	0.011	0.996	0.993–0.999
Emergency surgery	0.587	0.222	0.008	1.799	1.164–2.781
Cardiopulmonary bypass time (min)	0.006	0.002	< 0.001	1.006	1.003–1.010
Surgery with circulatory arrest	0.875	0.379	0.021	2.400	1.142–5.042
Rectal temperature when circulatory arrest (°C)	−0.129	0.054	0.017	0.879	0.792–0.977
Constant	−3.315	1.842	0.072	0.036	

eGFR, Estimated Glomerular Filtration Rate.

TABLE 3 | Features used to build machine learning prediction model.

Information on admission
Age (years)
Pulse (beats/min)
BMI (kg/m^2)
Diastolic pressure (mmHg)
Echocardiographic results
Left ventricular ejection fraction (%)
Preoperative laboratory examination results
Absolute value of leukocytes (10^9/L)
Platelets (10^9/L)
D-dimer (ng/mL)
INR
APTT (s)
ALT (U/mL)
Albumin (g/mL)
eGFR (ml/min/1.73 m^2)
Fasting blood glucose (mmol/L)
Surgery
Emergency surgery

BMI, body mass index; INR, international normalized ratio; APTT, activated partial thromboplastin time; ALT, alanine aminotransferase; eGFR, estimated glomerular filtration rate.

Leicester scoring system (AUC = 0.72, 95% CI: 0.68–0.77, $P < 0.001$) (**Table 4**).

External Validation

The external validation group included 319 patients. The comparison of the internal and external validation group characteristics is presented in **Table 5**. The median of the average age of patients in the external validation group was 50.0 (16.0) years. The incidence of ARF after aortic surgery was similar to that in the internal validation group (11.0 vs. 11.5%, $P = 0.807$).

After probability calibration, the Brier score of the machine learning prediction model using the external validation data was 0.087, which showed that the prediction model had good discrimination and calibration. Using external validation data for evaluation, we found that the XGBoost model after probability calibration (AUC = 0.81, 95% CI: 0.75–0.88) performed better than the LR prediction model (AUC = 0.75, 95% CI: 0.67–0.83, $P = 0.03$), the Cleveland scoring system (AUC = 0.71, 95% CI: 0.63–0.80, $P = 0.04$), the SRI scoring system (AUC = 0.70, 95% CI: 0.61–0.79, $P = 0.02$), and the Leicester scoring system (AUC = 0.67, 95% CI: 0.59–0.75, $P = 0.002$) (**Table 4**).

Finally, to make the XGBoost prediction model easy to use, we developed an application (https://ljzyal.github.io/ARF/) for clinical use. The application used a probability-calibrated XGBoost prediction model, which had the same performance as the prediction model in external validation. We set the cut-off value based on the results of external validation. The prediction model had a sensitivity of 82.9% and a specificity of 67.6%. The risk calculated by the application increased with the possibility of postoperative ARF.

DISCUSSION

Factors Influencing Postoperative ARF and the Role of a Prediction Model

ARF is the end stage of acute kidney injury, and it is the most common major complication following cardiac surgery (18).

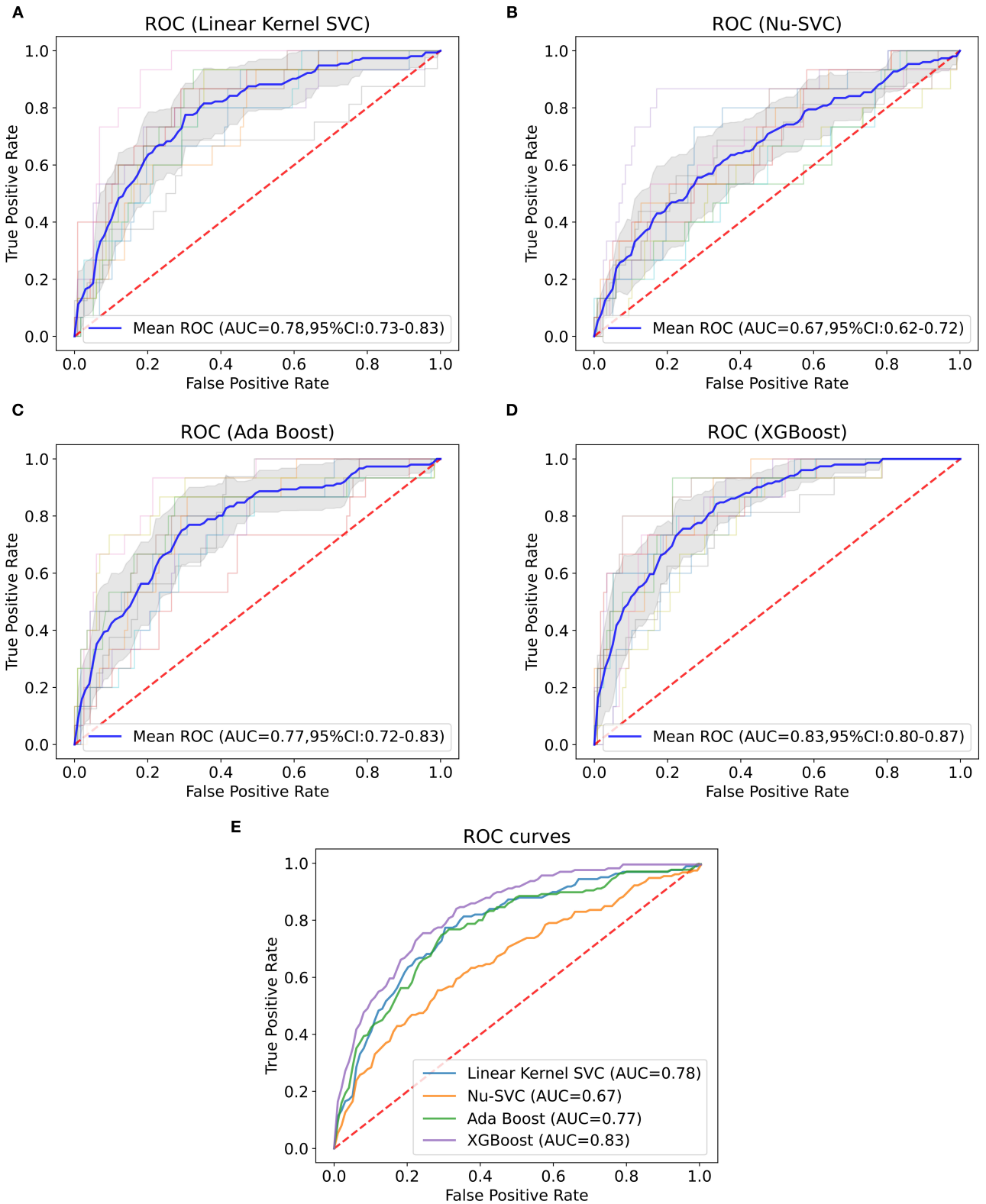


FIGURE 1 | Mean ROC curve and AUC of machine learning models. This figure depicts the mean ROC curve and AUC of the linear kernel SVC **(A)**, Nu-SVC **(B)**, AdaBoost **(C)** and XGBoost **(D)** using internal validation data (n = 1,318). The blue line represents the mean of each ROC curve after 10-fold cross-validation. The shaded area is the 95% confidence interval of the mean ROC curve. The other translucent lines are ROC curves for each cross-validation. **(E)** Comparison of the mean ROC curves for each algorithm.

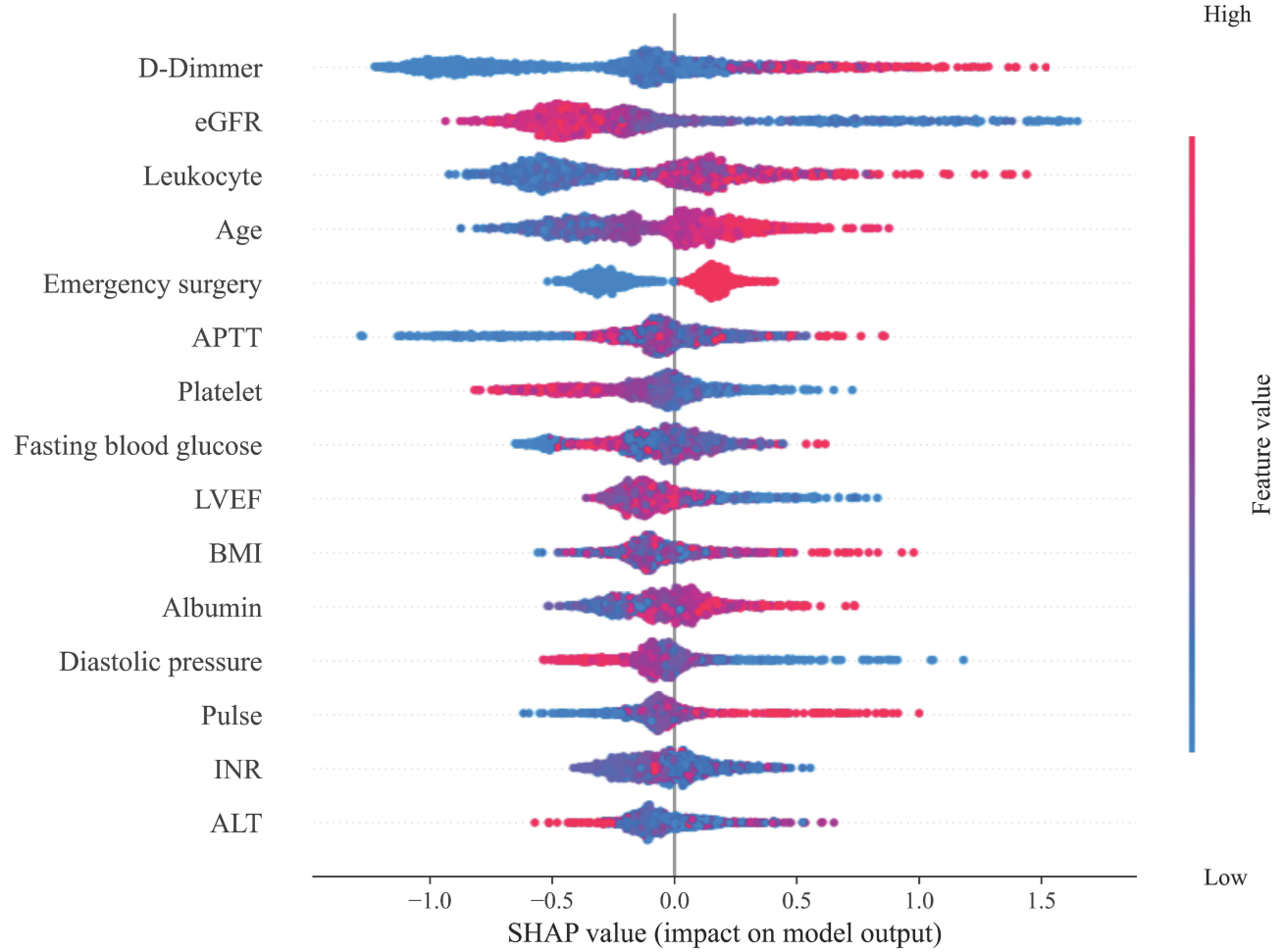


FIGURE 2 | Feature importance analysis. This figure shows the results of the analysis on the importance of the features in the XGBoost model through the SHAP method. Each feature value of each patient is marked as a dot on the graph. The color of the dot represents the degree of deviation of the feature value from the overall value according to the ordinate, and purple represents that the feature of the patient is close to the mean of the feature of the overall patient value. The SHAP value of the dot indicates the influence of the feature on the prediction result. A negative SHAP value indicates that the patient's risk of ARF is reduced, while a positive SHAP value indicates that the patient's risk of ARF is increased.

ARF can lead to poor patient prognosis and is independently associated with increased morbidity and mortality after cardiac surgery (19). In this study, the incidence of postoperative ARF in AAS patients reached 11.5%, and patients with ARF had a longer ICU length of stay, longer ventilator use time and a worse prognosis.

The mechanism of ARF after cardiac surgery remains to be elucidated, and its pathogenesis is currently thought to be related to renal hypoperfusion, tissue ischaemia-reperfusion injury and the inflammatory response (20). Previous studies have shown that risk factors for ARF include female sex, advanced age, previous heart surgery, chronic obstructive pulmonary disease, diabetes, complex heart surgery, prolonged cardiopulmonary bypass, rapid heart rate, emergency surgery, and intraoperative infusion of 2 or more packed red blood cell (RBC) units (21–23). In this study, we found that patients with AAS have more complex risk factors for postoperative ARF. In addition to the above factors, we found that postoperative ARF was also related to preoperative leukocyte and platelet counts, which may be because the occurrence of ARF is related to the inflammatory response (20). The higher preoperative leukocyte count may indicate that the inflammatory response caused by AAS is more serious. This effect may continue to play a role after surgery, making postoperative ARF more likely to occur. Furthermore, abnormal blood coagulation is another potential mechanism of ARF (24). Lower preoperative platelets may be a manifestation of a hypercoagulable state and intravascular coagulation, which makes AAS patients with higher preoperative platelet counts less prone to postoperative ARF.

Early identification of patients with a higher ARF risk can help clinicians strengthen patient monitoring and take measures to prevent ARF. Many studies have used risk factors or novel

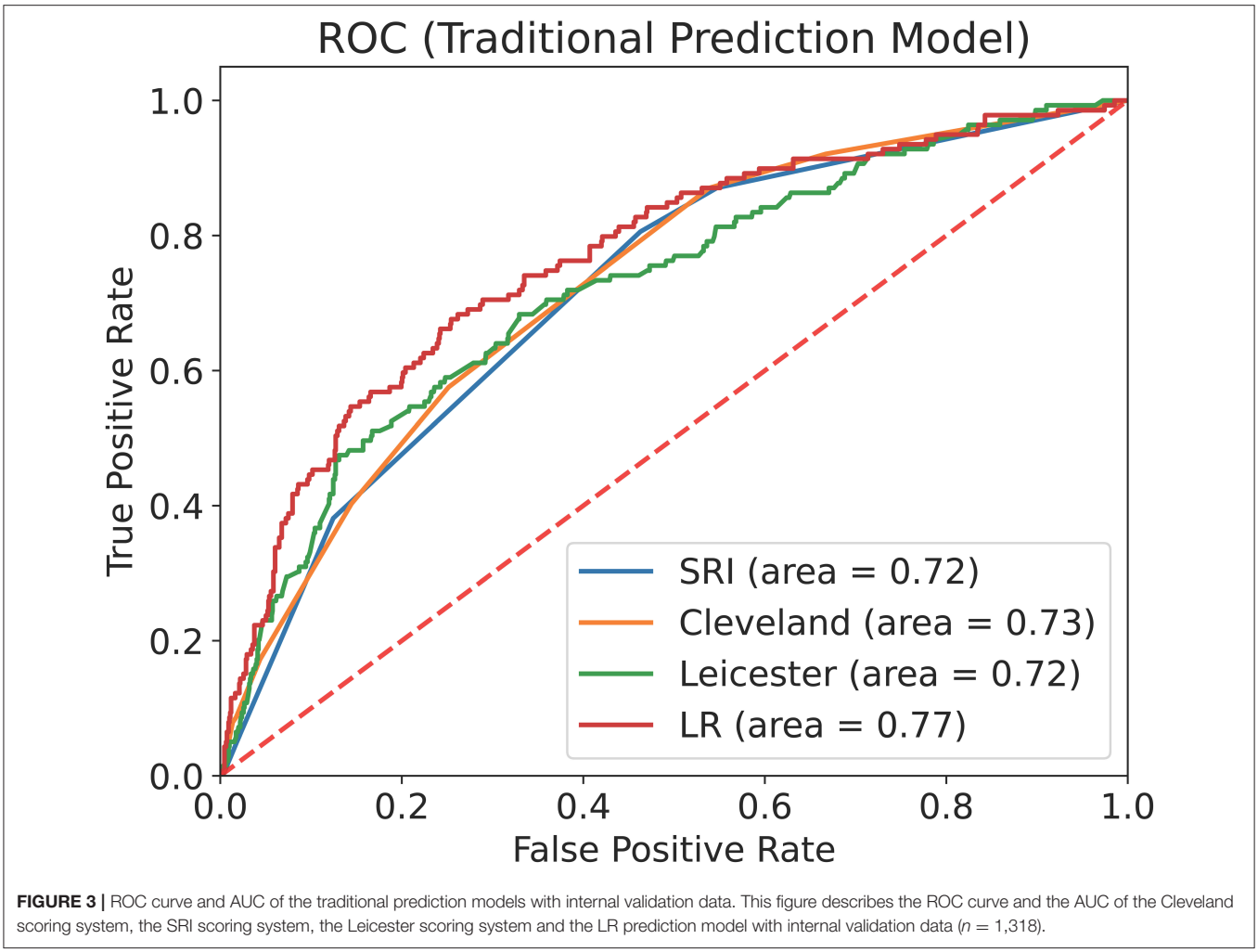


FIGURE 3 | ROC curve and AUC of the traditional prediction models with internal validation data. This figure describes the ROC curve and the AUC of the Cleveland scoring system, the SRI scoring system, the Leicester scoring system and the LR prediction model with internal validation data (n = 1,318).

biomarkers to build prediction models for ARF (3–6, 25). Novel biomarker-related prediction methods, however, are usually cumbersome, and no new biomarker has been widely accepted (25). Currently, the best-performing large-sample model is poor at predicting ARF after complex surgery (6). Aortic surgery usually results in a higher incidence of postoperative ARF; therefore, postoperative ARF prediction methods for complex heart and aortic surgeries are necessary.

Clinical Applications and Strategies

According to the results of this study, we recommend the following measures to reduce the occurrence of ARF. Once the model is used to predict the postoperative risk of ARF, for those with a low predicted risk, it is recommended to perform surgery in a timely fashion once the patient is surgically prepared.

For patients with a higher predicted risk of ARF, it is recommended to attempt to ameliorate the modifiable risk factors that are included in the prediction model. This could be achieved by improving preoperative preparations, such as administering antimicrobials, considering platelet transfusion, controlling blood glucose levels, maintaining adequate diastolic blood pressure, and controlling heart rate. In addition, during the operation, it is recommended to pay more attention to renal function and to maintain renal perfusion by taking measures to maintain circulatory stability. In addition to the aforementioned modifications, intraoperative innovations in surgical methods should be adopted, which can help reduce the operative and CPB time.

Additionally, for patients noted to be at a higher risk of ARF, it is also recommended to use stricter monitoring and more favorable preventive and treatment measures in perioperative management, which could include minimizing the use of nephrotoxic drugs and administering treatment for renal injury as soon as possible to prevent patients from progressing to ARF.

Features of Machine Learning Prediction Models

Risk prediction plays an important role in cardiovascular disease research. As the most commonly used traditional predictive model, LR sometimes cannot handle complex clinical data

TABLE 4 | Performance of machine learning prediction model and scoring system.

	Prediction methods	AUC	95% CI of AUC	*P value*
Internal validation				
	Machine learning prediction model (XGBoost)	0.82	0.79–0.85	
	LR prediction model	0.77	0.73–0.81	< 0.001
	Cleveland scoring system	0.73	0.69–0.77	< 0.001
	SRI scoring system	0.72	0.68–0.76	< 0.001
	Leicester scoring system	0.72	0.68–0.77	< 0.001
External validation				
	Machine learning prediction model (XGBoost)	0.81	0.75–0.88	
	LR prediction model	0.75	0.67–0.83	0.03
	Cleveland scoring system	0.71	0.63–0.80	0.04
	SRI scoring system	0.70	0.61–0.79	0.02
	Leicester scoring system	0.67	0.59–0.75	0.002

AUC, area under the receiver operating characteristic (ROC) curve; XGBoost, eXtreme Gradient Boosting; LR, logistic regression; SRI, simplified renal index.
The AUC of the ROC curve was compared with a machine learning prediction model (XGBoost), and DeLong's test was used to calculate the P value.

and thus cannot obtain an ideal predictive model. Conversely, machine learning can handle complex clinical data and thus potentially has more advantages (26). In this study, by selecting AAS as a disease process for focused research, we found that the performance of machine learning predictive models is better than that of traditional predictive models. This suggests that machine learning algorithms are more suitable for building clinical prediction models and have a higher performance than LR.

In this study, we found that the XGBoost algorithm has the best prediction performance and still has excellent performance in external validation. XGBoost is a machine learning algorithm that uses classification and regression trees as weak classifiers (14). Compared with other algorithms, the XGBoost algorithm allows easy adjustment of parameters and can deal with nonlinear features. It usually has higher sensitivity and specificity when overfitting is avoided. In most cases, XGBoost has higher prediction performance than other algorithms (27, 28).

Machine learning algorithms are also suitable for the construction of other predictive models, which have excellent performance and can be continuously trained with new data to have greater potential. After the prediction model is established, the newly collected data can be used to continue training, thereby enhancing the generalization ability of the prediction model.

Machine learning algorithms have been considered a black box in the past, which is the main disadvantage compared to LR. However, the SHAP method can explain the machine learning prediction model. We used the SHAP method to analyse the importance of particular features in the prediction model. This method can analyse the impact of each variable on each patient so the prediction model is interpretable. We found that the SHAP method is effective in determining the importance of particular individual features.

To compare with LR, all features included in an LR are also considered important features in machine learning prediction models. Concurrently, however, the SHAP method also judges other features, those that are not considered statistically significant in the LR, as important features. This possibly results in LR not including certain relevant features, whereas the SHAP method does not exclude such features. According to the results of the feature importance analysis, we can also judge the impact of each variable on the results to determine the patient's treatment direction to prevent postoperative ARF.

Limitations

First, all data for this study were sourced from China. Due to ethnic differences, the performance of our predictive model in other countries may decrease. However, our research method is innovative, and it is feasible to establish such a model in other countries through this investigational method. Second, machine learning algorithms are more complex than LR, and model representation is also very complicated (29). Our predictive model cannot be similar to LR, and it does not provide a scoring system for clinicians. We have therefore developed an online application for convenience. Third, our machine learning prediction model needs more extensive data for verification. Finally, although the performance of our prediction model was better than that of LR, some data were not involved in the initial data collection, such as detailed laboratory test results, detailed medical history and detailed documentation of the use of nephrotoxic drugs. Supervised machine learning can improve the model after supplementing these data, and consequently, our predictive model has the potential to improve.

In summary, our findings suggest that machine learning prediction models can provide better prediction performance than traditional LR prediction models and other existing risk scoring systems for AAS and complex cardiac and aortic surgeries. This predictive model is helpful for the early detection of patients with high ARF risk, thus enabling clinicians to take early measures to prevent and treat ARF.

TABLE 5 | Comparison of characteristics in the internal and external validation groups.

	Internal validation group	External validation group	*P* value
Number of patients (cases)	1,318	319	
Incidence of postoperative ARF	151 (11.5%)	35 (11.0%)	0.81
Need CRRT treatment (cases)	137 (90.7%)	30 (85.7%)	0.36
Information on admission			
Gender, female (cases)	301 (22.8%)	65 (20.4%)	0.34
Age (years)	50.0 (42.0–57.0)	50.0 (43.0–59.0)	0.10
Pulse (beats/min)	80.0 (75.0–85.0)	80.0 (74.0–85.0)	0.62
Height (cm)	170.0 (166.0–175.0)	170.0 (165.0–175.0)	0.59
Weight (kg)	75.0 (65.0–82.4)	74.0 (65.0–83.0)	0.42
Body mass index (kg/m^2)	25.4 (22.9–27.8)	25.4 (23.1–27.8)	0.71
Systolic pressure (mmHg)	130.0 (120.0–142.0)	135.0 (120.0–160.0)	< 0.001
Diastolic pressure (mmHg)	78.0 (70.0–84.1)	80.0 (70.0–95.0)	< 0.001
Medical history			
Smoking history (cases)	507 (38.5%)	96 (30.1%)	0.005
History of previous cardiac surgery (cases)	107 (8.1%)	20 (6.3%)	0.27
Peripheral vascular disease history (cases)	9 (0.7%)	3 (0.9%)	0.71
Echocardiographic results			
Left ventricular ejection fraction (%)	63.0 (60.0–66.0)	61.0 (57.0–66.0)	0.02
Preoperative laboratory examination results			
Absolute value of leukocytes (10^9/L)	9.69 (7.00–13.18)	11.08 (7.78–14.29)	< 0.001
Platelets (10^9/L)	185.0 (147.0–228.3)	184.0 (143.0–236.0)	0.93
Hemoglobin (g/L)	137.0 (122.0–148.0)	134.0 (122.0–147.0)	0.19
CK-MB (ng/mL)	1.88 (0.90–9.30)	3.30 (1.00–11.60)	< 0.001
Lactate dehydrogenase (U/L)	221.5 (179.0–288.3)	224.0 (189.0–277.4)	0.32
D-dimer (ng/mL)	1,100.0 (270.8–3,269.3)	715.9 (18.4–4,230.0)	< 0.001
INR	31.9 (28.9–36.5)	34.0 (29.1–39.7)	0.39
APTT (s)	48.8 (39.6–60.1)	50.4 (43.0–58.4)	< 0.001
Blood amylase (U/dL)	21.0 (15.0–34.0)	23.0 (15.0–36.0)	0.14
ALT (U/mL)	22.0 (18.0–32.0)	22.0 (18.0–33.0)	0.14
AST (U/mL)	39.1 (35.6–42.1)	38.5 (35.0–42.0)	0.56
Albumin (g/mL)	78.3 (64.6–99.6)	80.4 (66.8–101.6)	0.12
Creatinine (μmol/L)	6.30 (4.99–8.10)	6.40 (5.10–8.70)	0.48
BUN (mmol/mL)	95.0 (73.1–107.1)	92.3 (70.4–107.3)	0.05
eGFR (ml/min/1.73 m^2)	6.49 (5.39–7.77)	6.63 (5.40–7.65)	0.22
Fasting blood glucose (mmol/L)	9.69 (7.00–13.18)	11.08 (7.78–14.29)	0.94
Diagnosis			
Coronary artery disease (cases)	35 (2.7%)	6 (1.9%)	0.43
Congestive heart failure (cases)	26 (2.0%)	1 (0.3%)	0.04
Chronic respiratory disease (cases)	31 (2.4%)	15 (4.7%)	0.02
Hypertension (cases)	905 (68.7%)	216 (67.7%)	0.74
Diabetes (cases)	62 (4.7%)	13 (4.1%)	0.63
Surgery			
Operative duration (min)	405.0 (340.0–479.0)	390.0 (321.2–480.0)	0.21
Emergency surgery (cases)	650 (49.3%)	164 (51.4%)	0.50
Cardiopulmonary bypass time (min)	186.0 (144.0–224.0)	184.0 (135.2–236.0)	0.70
Aortic cross-clamp time (min)	104.0 (82.6–131.0)	108.0 (82.0–140.0)	0.13
With circulatory arrest (cases)	997 (75.6%)	223 (69.9%)	0.04
Circulatory arrest time (min)	21.0 (17.4–27.0)	21.2 (17.0–28.0)	0.63
Nasopharyngeal temperature when circulatory arrest (°C)	24.1 (23.1–24.9)	24.2 (23.0–25.0)	0.32
Rectal temperature when circulatory arrest (°C)	25.6 (24.8–26.7)	25.8 (24.5–27.0)	0.84
RBC transfusion volume (U)	4.00 (0.00–6.00)	4.00 (2.00–6.00)	0.006

CRRT, continuous renal replacement therapy; INR, international normalized ratio; APTT, activated partial thromboplastin time; ALT, alanine aminotransferase; AST, aspartate transaminase; BUN, blood urea nitrogen; eGFR, estimated glomerular filtration rate.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Beijing Anzhen Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

HZ and WJ designed the research. JL and MG analyzed the data and wrote the paper. YJ analyzed the data and revised the manuscript. LS, LH, RF, TG, ZZ, CZ, GZ, XQ, CQ, and YC were responsible for the data collection. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Key Research and Development Program of China (2017YFC1308000), Capital Health Development Research Project (2018-2-2066 and 2018-4-2068), National Science Foundation of China (81600362 and 81800404), Beijing Lab for Cardiovascular Precision Medicine (PXM2017_014226_000037), Beijing Advanced Innovation Center for Big Data-based Precision Medicine (PXM2021_014226_000026), Beijing Municipal Administration of Hospitals' Youth Program (QML20180601), and Foundation of Beijing Outstanding Young Talent Training Program (2017000021469G254). The funders did not have a role in the study design, collection, analysis or reporting of data, preparation of the manuscript or decision to submit for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.728521/full#supplementary-material

REFERENCES

1. Erbel R, Aboyans V, Boileau C, Bossone E, Bartolomeo RD, Eggebrecht H, et al. 2014 ESC Guidelines on the diagnosis and treatment of aortic diseases: Document covering acute and chronic aortic diseases of the thoracic and abdominal aorta of the adult. The Task Force for the Diagnosis and Treatment of Aortic Diseases of the European Society of Cardiology (ESC). *Eur Heart J.* (2014) 35:2873–926. doi: 10.1093/eurheartj/ehu281
2. Englberger L, Suri RM, Greason KL, Burkhart HM, Sundt TM. 3rd, Daly RC, et al. Deep hypothermic circulatory arrest is not a risk factor for acute kidney injury in thoracic aortic surgery. *J Thorac Cardiovasc Surg.* (2011) 141:552–8. doi: 10.1016/j.jtcvs.2010.02.045
3. Thakar CV, Arrigain S, Worley S, Yared JP, Paganini EP, A. clinical score to predict acute renal failure after cardiac surgery. *J Am Soc Nephrol.* (2005) 16:162–8. doi: 10.1681/ASN.2004040331
4. Mehta RH, Grab JD, O'Brien SM, Bridges CR, Gammie JS, Haan CK, et al. Bedside tool for predicting the risk of postoperative dialysis in patients undergoing cardiac surgery. *Circulation.* (2006) 114:2208–16. doi: 10.1161/CIRCULATIONAHA.106.635573
5. Wijeysundera DN, Karkouti K, Dupuis JY, Rao V, Chan CT, Granton JT, et al. Derivation and validation of a simplified predictive index for renal replacement therapy after cardiac surgery. *JAMA.* (2007) 297:1801–9. doi: 10.1001/jama.297.16.1801
6. Birnie K, Verheyden V, Pagano D, Bhabra M, Tilling K, Sterne JA, et al. Predictive models for kidney disease: improving global outcomes (KDIGO) defined acute kidney injury in UK cardiac surgery. *Crit Care.* (2014) 18:606. doi: 10.1186/s13054-014-0606-x
7. Collins GS, Reitsma JB, Altman DG, Moons KG, Group T. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. The TRIPOD group. *Circulation.* (2015) 131:211–9. doi: 10.1161/CIRCULATIONAHA.114.014508
8. Kellum JA, Lameire N, Group KAGW. Diagnosis, evaluation, and management of acute kidney injury: a KDIGO summary (Part 1). *Crit Care.* (2013) 17:204. doi: 10.1186/cc11454
9. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF. 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* (2009) 150:604–12. doi: 10.7326/0003-4819-150-9-200905050-00006
10. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, et al. Missing value estimation methods for DNA microarrays. *Bioinformatics.* (2001) 17:520–5. doi: 10.1093/bioinformatics/17.6.520
11. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* (1988) 44:837–45. doi: 10.2307/2531595
12. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* (2011) 12:2825–30.
13. Freund Y, Schapire RE, editors. A desicion-theoretic generalization of on-line learning and an application to boosting. *European conference on computational learning theory*. Springer. (1995) doi: 10.1007/3-540-59119-2_166
14. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, California, USA: Association for Computing Machinery (2016). p. 785–94. doi: 10.1145/2939672.2939785
15. Lundberg SM, Lee S-I, A. Unified Approach to Interpreting Model Predictions. *Adv Neural Inf Process Syst.* (2017) 30:4765–74.
16. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. *Eur Heart J.* (2017) 38:500–7. doi: 10.1093/eurheartj/ehw188
17. Platt JC. *Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods*. Advances in Large Margin Classifiers. (1999).
18. Bove T, Monaco F, Covello RD, Zangrillo A. Acute renal failure and cardiac surgery. *HSR Proc Intensive Care Cardiovasc Anesth.* (2009) 1:13–21.
19. Ortega-Loubon C, Fernandez-Molina M, Carrascal-Hinojal Y, Fulquet-Carreras E. Cardiac surgery-associated acute kidney injury. *Ann Card Anaesth.* (2016) 19:687–98. doi: 10.4103/0971-9784.191578
20. Chew STH, Hwang NC. Acute Kidney Injury After Cardiac Surgery: A Narrative Review of the Literature. *J Cardiothorac Vasc Anesth.* (2019) 33:1122–38. doi: 10.1053/j.jvca.2018.08.003
21. Pacini D, Pantaleo A, Di Marco L, Leone A, Barberio G, Parolari A, et al. Risk factors for acute kidney injury after surgery of the thoracic aorta using antegrade selective cerebral perfusion and moderate hypothermia. *J Thorac Cardiovasc Surg.* (2015) 150:127–33 e1. doi: 10.1016/j.jtcvs.2015.04.008
22. Khan UA, Coca SG, Hong K, Koyner JL, Garg AX, Passik CS, et al. Blood transfusions are associated with urinary biomarkers of kidney injury in cardiac surgery. *J Thorac Cardiovasc Surg.* (2014) 148:726–32. doi: 10.1016/j.jtcvs.2013.09.080

23. Shiao CC, Huang YT, Lai TS, Huang TM, Wang JJ, Huang CT, et al. Perioperative body weight change is associated with in-hospital mortality in cardiac surgical patients with postoperative acute kidney injury. *PLoS ONE.* (2017) 12:e0187280. doi: 10.1371/journal.pone.0187280
24. Nguyen TC, Cruz MA, Carcillo JA. Thrombocytopenia-associated multiple organ failure and acute kidney injury. *Crit Care Clin.* (2015) 31:661–74. doi: 10.1016/j.ccc.2015.06.004
25. Wang Y, Bellomo R. Cardiac surgery-associated acute kidney injury: risk factors, pathophysiology and treatment. *Nat Rev Nephrol.* (2017) 13:697–711. doi: 10.1038/nrneph.2017.119
26. Hamet P, Tremblay J. Artificial intelligence in medicine. *Metabolism.* (2017) 69S:S36–S40. doi: 10.1016/j.metabol.2017.01.011
27. Ruan Y, Bellot A, Moysova Z, Tan GD, Lumb A, Davies J, et al. Predicting the risk of inpatient hypoglycemia with machine learning using electronic health records. *Diabetes Care.* (2020) 43:1504–11. doi: 10.2337/dc19-1743
28. Leung WK, Cheung KS Li B, Law SYK, Lui TKL. Applications of machine learning models in the prediction of gastric cancer risk in patients after Helicobacter pylori eradication. *Aliment Pharmacol Ther.* (2021) 53:864–72. doi: 10.1111/apt.16272
29. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: applying machine learning to address analytic challenges. *Eur Heart J.* (2017) 38:1805–14. doi: 10.1093/eurheartj/ehw302

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

ORIGINAL RESEARCH
published: 18 January 2022
doi: 10.3389/fmed.2021.814566

A Simple Weaning Model Based on Interpretable Machine Learning Algorithm for Patients With Sepsis: A Research of MIMIC-IV and eICU Databases

*Wanjun Liu[1,2], Gan Tao[1], Yijun Zhang[1,2], Wenyan Xiao[1,2], Jin Zhang[1,2], Yu Liu[3], Zongqing Lu[1,2], Tianfeng Hua[1,2] and Min Yang[1,2]**

[1] The 2nd Department of Intensive Care Unit, The Second Affiliated Hospital of Anhui Medical University, Hefei, China, [2] The Laboratory of Cardiopulmonary Resuscitation and Critical Care Medicine, The Second Affiliated Hospital of Anhui Medical University, Hefei, China, [3] Key Laboratory of Intelligent Computing and Signal Processing, Ministry of Education, Anhui University, Hefei, China

Background: Invasive mechanical ventilation plays an important role in the prognosis of patients with sepsis. However, there are, currently, no tools specifically designed to assess weaning from invasive mechanical ventilation in patients with sepsis. The aim of our study was to develop a practical model to predict weaning in patients with sepsis.

Methods: We extracted patient information from the Medical Information Mart for Intensive Care Database-IV (MIMIC-IV) and the eICU Collaborative Research Database (eICU-CRD). Kaplan–Meier curves were plotted to compare the 28-day mortality between patients who successfully weaned and those who failed to wean. Subsequently, MIMIC-IV was divided into a training set and an internal verification set, and the eICU-CRD was designated as the external verification set. We selected the best model to simplify the internal and external validation sets based on the performance of the model.

Results: A total of 5020 and 7081 sepsis patients with invasive mechanical ventilation in MIMIC-IV and eICU-CRD were included, respectively. After matching, weaning was independently associated with 28-day mortality and length of ICU stay ($p < 0.001$ and $p = 0.002$, respectively). After comparison, 35 clinical variables were extracted to build weaning models. XGBoost performed the best discrimination among the models in the internal and external validation sets (AUROC: 0.80 and 0.86, respectively). Finally, a simplified model was developed based on XGBoost, which included only four variables. The simplified model also had good predictive performance (AUROC:0.75 and 0.78 in internal and external validation sets, respectively) and was developed into a web-based tool for further review.

Conclusions: Weaning success is independently related to short-term mortality in patients with sepsis. The simplified model based on the XGBoost algorithm provides good predictive performance and great clinical applicablity for weaning, and a web-based tool was developed for better clinical application.

Keywords: sepsis, invasive mechanical ventilation, weaning, XGBoost, simple prediction model

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Songqiao Liu,
Southeast University, China
Kay Choong See,
National University Hospital, Singapore

***Correspondence:**
Min Yang
512130761@qq.com

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *13 November 2021*
Accepted: *13 December 2021*
Published: *18 January 2022*

Citation:
Liu W, Tao G, Zhang Y, Xiao W, Zhang J, Liu Y, Lu Z, Hua T and Yang M (2022) A Simple Weaning Model Based on Interpretable Machine Learning Algorithm for Patients With Sepsis: A Research of MIMIC-IV and eICU Databases. Front. Med. 8:814566. doi: 10.3389/fmed.2021.814566

INTRODUCTION

Difficult weaning or prolonged invasive mechanical ventilation is more common in patients with sepsis (1, 2). Lung susceptibility to ventilatory injury is thought to be increased by sepsis (3, 4), and mechanical ventilation may also lead to the exacerbation of pulmonary infection (5). Prolonged mechanical ventilation can lead to a poor prognosis (6, 7). However, insufficient duration of mechanical ventilation is unfavorable for patients. Weaning in unprepared patients leads to increased mortality and prolonged ICU stay (8). Therefore, the choice of an appropriate weaning time is of great importance.

Previous studies on weaning have evaluated numerous methods on weaning, such as rapid shallow breathing index (RSBI) (9), spontaneous breathing experiment (SBT), and compliance, oxygenation, respiratory rate, and pressure (CROP) index. Nevertheless, weaning factors specific to patients with sepsis are scarce. Unfortunately, the accuracy of these factors in predicting weaning is unsatisfactory (10, 11). Moreover, weaning from mechanical ventilation has also been shown to be related to consciousness, diaphragmatic function, and cardiac function (12–14). Traditional prediction of weaning has several limitations. On the one hand, traditional methods of weaning as a complex process are inadequate for the use of clinical indicators utilization. On the other hand, traditional methods have deficits in predictive performance due to disease-related differences in the target population, as there are no specific target populations.

Given the rapid development of clinical medicine, a refined weaning scheme is needed to meet the demands of clinical development. A simple and reliable weaning program could not only effectively assist clinicians but also improve the patient prognosis, especially in patients with sepsis with ventilator dependence.

In this study, we aimed to develop a reliable model for predicting weaning success in patients with sepsis. To this end, we extracted data within 24 h from patients with sepsis before weaning from a large dataset. Features were selected based on their clinical availability and explained by their importance. In addition, our model was further validated using datasets from various sources.

METHODS

Data Source

Our study was a retrospective cohort study based on the MIMIC-IV (version 1.0) database. This database contains over 40,000 ICU patients from Beth Israel Deaconess Medical Center between 2008 and 2019. Moreover, we used an independent external validation set called eICU-CRD Collaborative Research (eICU-CRD) Database (version 2.0), which is a multicenter database of over 200,000 ICU admissions in the United States. We carefully studied the courses and obtained permission to use the database (record ID 39691989). Because the patient privacy information was encrypted in the database, the ethics committee at the two medical centers did not require informed consent.

Patients and Definitions

In this study, sepsis was diagnosed based on the Sepsis-3 criteria [(15); SOFA score $\geq$ 2, and suspicious infection]. The contents of a recent guideline had been considered before implementing the criteria for successful weaning (16). The definition of weaning success (WS) was as follows: (a) no intubation or invasive ventilation within 48 h after weaning, (b) no death within 48 h after weaning, and (c) noninvasive ventilation time was shorter than 48 h after weaning. The patients who experienced invasive mechanical ventilation and met the criteria for Sepsis-3 were included in both datasets. The exclusion criteria were as follows: (a) repeated ICU admissions and (b) age $<$ 18 years (**Figure 1**).

Data Collection and Variable Extraction

For modeling, we preferred variables with a single measurement. Clinical importance and shapley additive explanations (SHAP) values were used to further reduce the amount of data. Patient demographics were collected, including age, sex, and body mass index (BMI). Clinical and chemical variables had been extracted within 24 h prior to weaning to create models. The extracted variables were the worst value of the day, which was as follows: arterial blood gas [pH value, arterial oxygen partial pressure (PaO_2), arterial carbon dioxide partial pressure ($PaCO_2$), base excess (BE)], full blood count [white blood cell count (WBC), hemoglobin (HB), platelet (PLT)], laboratory index (creatinine, anion gap), vital signs [heart rate, respiratory rate, mean arterial pressure (MAP), peripheral oxygen saturation (SPO_2), temperature, oxygenation index (OI)], and urine output. In addition, data on therapeutic measures [days of invasive ventilation, days of antibiotic use, days of continuous renal replacement therapy (CRRT), and vasopressor therapy within 24 h]. The variables in the matched data were from the first day of ICU admission and included the variables listed above (**Supplementary Table 1**). To further balance the differences in baseline data between the patients with weaning failure (WF) and patients with weaning success, comprehensive indicators were extracted, such as the sequential organ failure assessment (SOFA) score, Glasgow coma scale (GCS) score, and simplified acute physiology score (SAPS II). Comorbidities, as well as infection classification (**Supplementary Table 4**), were also considered based on the recorded International Classification of Diseases codes (ICD-9 and ICD-10), and the Charlson comorbidity index was also calculated.

Data Analysis

Continuous variables are described as median and interquartile range (IQR). The Mann–Whitney U-test was used for statistical comparison between the two groups. Categorical variables were described as total number and percentage, and the chi-square test or Fisher's exact test was used for comparison between groups. Propensity score matching (PSM) was used to balance the differences between successful and unsuccessful weaning groups. Inverse probability weighting (IPW) (17) was used to further adjust for possible imbalances between the variables of the two groups. The Kaplan–Meier (K–M) curve was used to describe the 28-day survival rate between the two groups, and

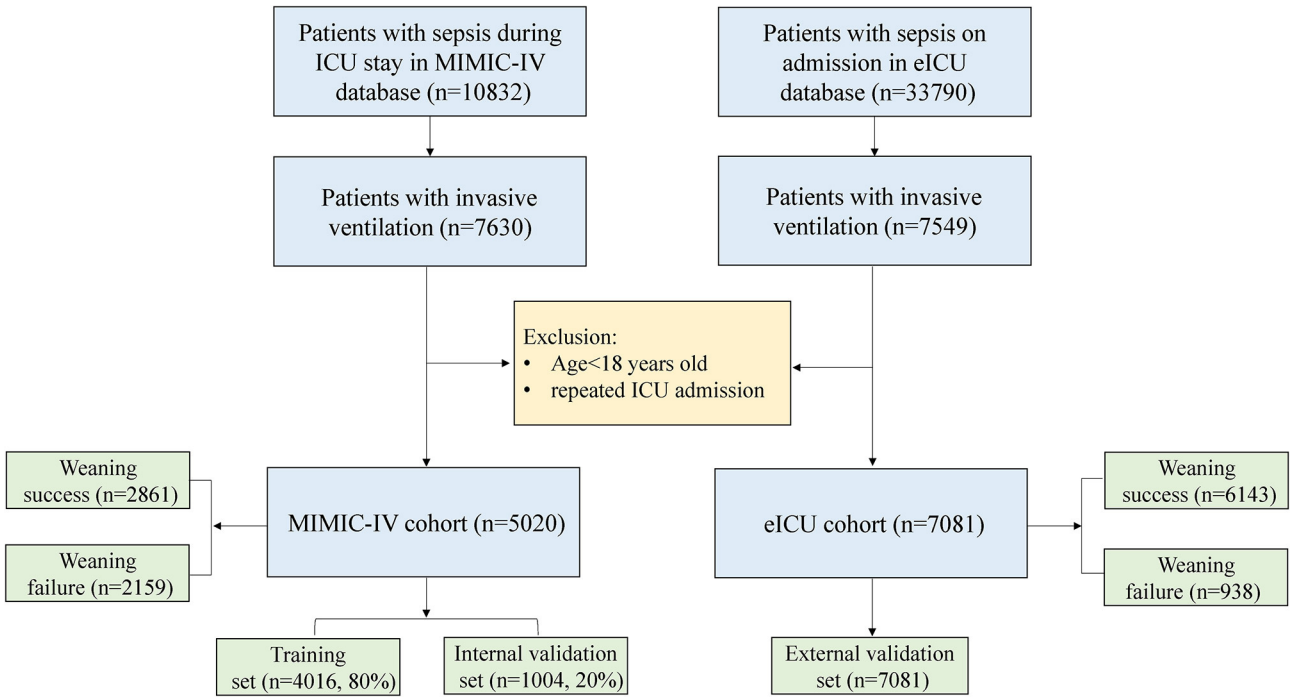


FIGURE 1 | The flow chart of data extraction. eICU-CRD, eICU-CRD Collaborative Research Database. MIMIC-IV, Medical Information Mart for Intensive Care-IV; ICU, intensive care unit.

the differences in survival rates between groups were compared using the log-rank test.

After comparison, we divided the MIMIC-IV data into two parts: 80% as a training set and 20% as an internal validation set, and the integrated machine-learning algorithm eXtremely Gradient Boosting (XGBoost) to construct a weaning prediction model, which is based on multiple decision trees with gradient boost as a learning framework. The hyperparameters were optimized using a grid search (**Supplementary Table 3**). Other models, such as KNearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR), were also derived from the training set and applied to the test set. The prediction efficiency of the models was compared using a receiver operating characteristic (ROC) curve. In addition, the model was further explained by the SHAP value, demonstrating a linear relationship through local weighted regression scatter smoothing (LOWESS).

Variables with more than 50% missing data were excluded. The missing features of the matched data and model data are shown in **Supplementary Figure 1A** and **Supplementary Table 2**. Missing values were input using the multiple imputation method. Due to the different missing datasets, we had extracted only the CROP and RSBI 24 h before weaning in MIMIC-IV, and the predictive performance was also compared with XGBoost in the internal validation set. In addition, the model was further simplified using the recursive feature elimination algorithm (RFE).

Structured query language was used to extract the data from these two databases. All statistical analyses were performed using R 3.6.2 (Chicago, Illinois) and Python (version 3.6.6), and statistical significance was set at $p < 0.05$.

RESULTS

Matching Baseline and Clinical Outcomes

As shown in **Figure 1**, 5,020 patients with sepsis who received invasive mechanical ventilation were ultimately included in the MIMIC-IV database. The baseline characteristics on the first day of ICU admission are shown in **Supplementary Table 1**. PSM and IPW were used to better balance the differences between the two groups. A total of 1,676 patients were included (**Supplementary Table 1**), and the standardized mean difference (SMD) between groups was significantly reduced (**Supplementary Figure 1B**). In comparison, there was a significant difference in the length of stay in the ICU between the two groups (**Figure 2B**, $p = 0.002$). Similarly, the WS group had a significantly lower 28-day mortality rate (**Figure 2A**, $p < 0.001$).

Baseline Characteristics of Models

In the MIMIC-IV cohort, successful weaning was associated with reductions in highest WBC, highest creatinine, highest anion gap, highest heart rate, highest respiratory rate, highest body temperature, highest PEEP level, antibiotic duration, invasive mechanical ventilation (IMV) duration, and vasopressor use 1 day before weaning (**Table 1**). Consequently, the performance of these indicators behaved similarly in the eICU-CRD cohort, except for age and the highest FiO_2 (**Table 1**). Regarding comorbidity, it was observed that successful weaning benefited

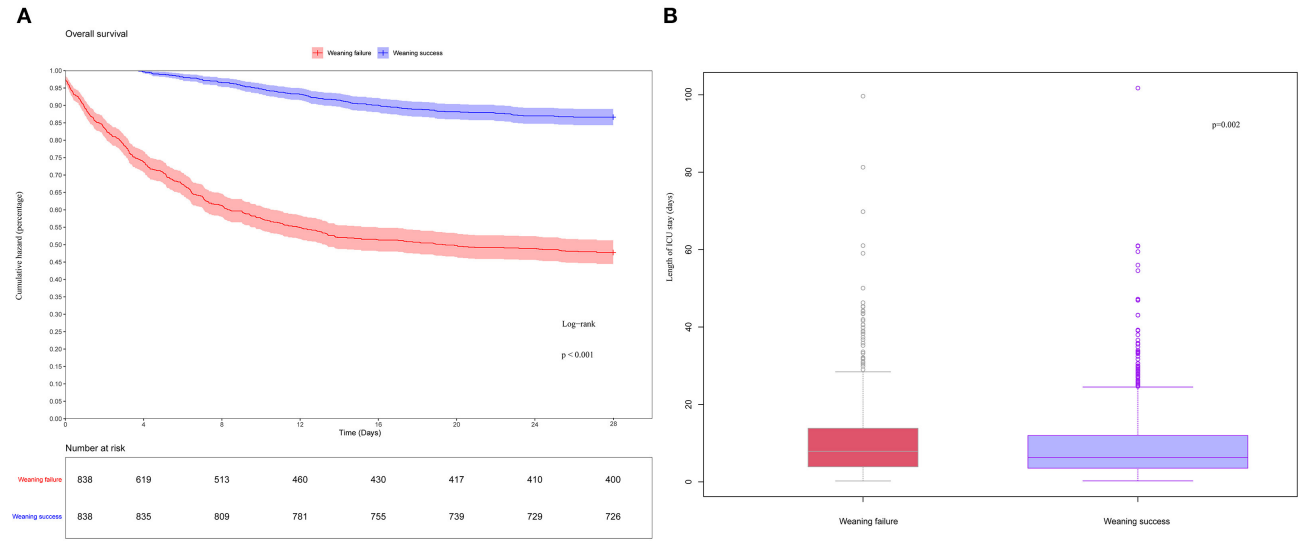


FIGURE 2 | (A) K-M curves estimated 28-day survival probability of weaning failure and weaning success patients. **(B)** Box-plot of weaning failure and weaning success patients.

chronic pulmonary disease, congestive heart failure, renal disease, and diabetes in the MIMIC-IV cohort. However, except for severe liver disease, other comorbidities were inconsistent between the WS and WF groups (**Table 1**).

Comparison and Explanation of Models

We trained the models using the training set from MIMIC-IV. As shown in **Table 2**, the XGBoost model with all available variables had a striking AUROC of 0.80 [95% confidence interval (CI): 0.77–0.82 in the internal validation set, and 0.86 (95% CI: 0.85–0.87)] in the external validation set, while the other five representative models had the highest AUROC, 00.74 (95% CI: 0.71–0.77) in the internal validation set, and 0.83 (95% CI: 0.82–0.84) in the external validation set. The final hyperparameter settings for XGBoost are listed in **Supplementary Table 3**. The SHAP values for the XGBoost model were assessed and are shown in **Figure 3A**. The importance of the variables was sorted by the gap value and is shown in **Figure 3B**. **Figures 4A,B** show the comparison between the XGBoost model and the other five models or predictive factors. As can be seen, the XGBoost model significantly outperformed the other five models or predictive factors in both the internal validation and external validation. Due to the extensive missing data, we did not show the ROC curves of CROP and RSBI in the external validation set. In addition, we performed a decision curve analysis (**Figure 4C**) and a calibration plot (**Figure 4D**) to illustrate the performance of the XGBoost model.

SHAP Values Depending on Variables

The probability of successful weaning increases with an increase in the following indicators: urine output, lowest base excess, GCS, lowest SPO_2, congestive heart failure, lowest pH, lowest map, highest $PaCO_2$, renal disease, lowest platelet count, BMI, and OI (**Figure 3A**). The contribution of each feature in the internal validation set is shown in **Figure 3B** in order of importance. Finally, we performed a partial dependency plot of the four contributing continuous variables to explain the impact of the change in value of each variable on the patients with WS, as shown in **Figure 5**. The remaining variables are shown in **Supplementary Figures 2**, **3**. As shown in the partial dependency plot, feature values are indicated by a blue scatter plot, with the linear relationship represented by a red curve, where SHAP values represent an increase in the probability of WS when the value is positive and *vice versa*.

Figure 5 shows the change trend of WS probability with the change in variables. For some variables, there is a discernable trend where the WS probability increases with an increase in the value of variables. These variables are urine output (**Figure 5C**), lowest BE (**Figure 5D**), GCS (**Supplementary Figure 2B**), and lowest SPO_2 (**Supplementary Figure 2D**). In contrast, the decrease in some indicators, such as the highest PEEP level (**Figure 5B**), the highest anion gap level (**Supplementary Figure 2A**), and age (**Supplementary Figure 2H**), suggests a decrease in WS probability. Additionally, some variables seem to have a reasonable value, and the best cut-off value for these variables can be roughly judged by the LOWESS curve (**Figure 5A**; **Supplementary Figures 2C,E–G**). To obtain the best WS possibility, the values of these variables should be kept close to the cut-off value.

The Web-Based Tool and an Example Scenario

We simplified the previous XGBoost model according to the variable importance. The four variables (IMV duration, highest PEEP level, urine output, and lowest base excess) with the highest importance were used to develop a simplified

TABLE 1 | Baseline characteristics of the MIMIC-IV and eICU cohorts.

Variables	MIMIC-IV cohort				eICU cohort			
	Overall (*n* = 5,020)	**Weaning failure (*n* = 2,159)**	**Weaning success (*n* = 2,861)**	***p***	**Overall (*n* = 7,081)**	**Weaning failure (*n* = 938)**	**Weaning success (*n* = 6,143)**	***p***
n	5,020	2,159	2,861		7,081	938	6,143	
Age (years)	68 (57, 78)	68 (56, 79)	68 (57, 78)	0.335	64 (53, 74)	68 (57, 77)	63 (52, 73)	<0.001
Male	2,944 (58.6)	1,241 (57.5)	1,703 (59.5)	0.154	3,976 (56.2)	511 (54.5)	3,465 (56.4)	0.283
BMI (kg/m^2)	28 (24, 33)	28 (24, 33)	28 (24, 32)	0.428	28 (23, 33)	27 (23, 33)	28 (23, 33)	0.02
Chronic pulmonary disease (*n*, %)	1,719 (34.2)	689 (31.9)	1,030 (36.0)	0.003	1,334 (18.8)	175 (18.7)	1,159 (18.9)	0.914
Congestive heart failure (*n*, %)	2,155 (42.9)	788 (36.5)	1,367 (47.8)	<0.001	1,174 (16.6)	158 (16.8)	1,016 (16.5)	0.852
Dementia (*n*, %)	239 (4.8)	88 (4.1)	151 (5.3)	0.056	207 (2.9)	40 (4.3)	167 (2.7)	0.012
Severe liver disease (*n*, %)	547 (10.9)	286 (13.2)	261 (9.1)	<0.001	209 (3.0)	60 (6.4)	149 (2.4)	<0.001
Renal disease (*n*, %)	1,463 (29.1)	574 (26.6)	889 (31.1)	0.001	949 (13.4)	151 (16.1)	798 (13.0)	0.011
Rheumatic disease (*n*, %)	211 (4.2)	84 (3.9)	127 (4.4)	0.375	165 (2.3)	21 (2.2)	144 (2.3)	0.934
Diabetes (%)	1,665 (33.2)	675 (31.3)	990 (34.6)	0.014	2,129 (30.1)	283 (30.2)	1,846 (30.1)	0.971
Charlson comorbidity index	6 (4, 8)	6 (4, 8)	6 (4, 8)	0.082	3 (2, 5)	4 (3, 6)	3.00 (2, 5)	<0.001
GCS	14 (10, 15)	13 (7, 15)	14 (10, 15)	<0.001	8 (6, 10)	4 (3, 8)	9 (6, 10)	<0.001
Highest WBC ($\times 10^9$/L)	14.1 (9.8, 19.9)	14.7 (9.8, 21.3)	13.8 (9.9, 18.9)	<0.001	11.9 (8.7, 16.4)	14.7 (9.7, 20.5)	11.7 (8.6, 15.8)	<0.001
Lowest hemoglobin (g/L)	9.2 (8.0, 10.6)	9.1 (7.9, 10.6)	9.3 (8.1, 10.6)	0.002	9.7 (8.4, 11.3)	9.3 (7.9, 10.8)	9.8 (8.5, 11.4)	<0.001
Lowest platelets ($\times 10^9$/L)	143 (89, 213)	136 (75, 212)	147 (100, 215)	<0.001	167 (113, 233)	141 (76, 212)	170 (118, 236)	<0.001
Highest creatinine (mg/dL)	1.3 (0.9, 2.3)	1.6 (1.0, 2.7)	1.2 (0.8, 1.9)	<0.001	1.1 (0.7, 1.8)	1.7 (1.0, 3.0)	1.0 (0.7, 1.6)	<0.001
Highest anion gap (mEq/L)	15.0 (13.0, 19.0)	17.0 (14.0, 22.0)	14.0 (12.0, 17.0)	<0.001	10.0 (8.0, 14.0)	13.0 (10.0, 18.0)	10.0 (7.7, 13.0)	<0.001
Lowest pH level	7.3 (7.3, 7.4)	7.3 (7.2, 7.4)	7.3 (7.3, 7.4)	<0.001	7.4 (7.3, 7.4)	7.3 (7.2, 7.4)	7.4 (7.3, 7.4)	<0.001
Lowest PaO_2 (mmHg)	80 (53, 106)	75 (48, 99)	84 (60, 110)	<0.001	85 (69, 115)	78.00 (61, 102)	86 (69, 117)	<0.001
Highest $PaCO_2$ (mmHg)	45 (39, 51)	44 (38, 52)	45 (40, 50)	0.837	43 (37, 49)	42 (36, 52)	43 (37, 49)	0.977
Lowest base excess (mEq/L)	−3.0 (−7.0, 0.0)	−4.0 (−10.0, 0.0)	−2.0 (−5.0, 0.0)	<0.001	−1.0 (−5.4, 2.0)	−5.5 (−12.4,0.6)	−0.6 (−5.0, 2.2)	<0.001
Highest heart rate (/min)	103 (90, 119)	108 (94, 124)	100 (88, 115)	<0.001	103 (90, 118)	112 (97, 129)	102.00 (89, 116)	<0.001
Highest respiratory rate (/min)	27 (23, 31)	29 (24, 33)	26 (22, 30)	<0.001	25 (21, 31)	30.00 (25, 35)	25.00 (21, 30)	<0.001
Lowest MAP (mmHg)	60 (54, 65)	59 (52, 64)	60 (56, 66)	<0.001	65 (57, 73)	59 (49, 68)	66 (59, 74)	<0.001
Highest body temperature (°C)	37.4 (37.0, 38.1)	37.5 (36.9, 38.2)	37.4 (37.1, 37.9)	0.021	37.4 (37.0, 37.9)	37.4 (36.9, 38.1)	37.4 (37.1, 37.9)	0.362
Lowest SPO_2	94 (91, 96)	93 (89, 95)	94 (92, 97)	<0.001	94 (91, 97)	91 (82, 94)	94 (91, 97)	<0.001
Highest PEEP (cmH_2O)	7 (5, 10)	9 (5, 12)	6 (5, 10)	<0.001	5 (5, 8)	5 (5, 10)	5.00 (5, 6)	<0.001
Lowest tidal volume (ml)	397 (328, 455)	395 (325 452)	398 (329 459)	0.154	422 (343, 497)	423.50 (356, 493)	422.00 (340, 498)	0.591
Lowest OI	174 (105, 250)	152 (91, 232)	188 (120, 260)	<0.001	206 (144, 282)	161 (98, 227)	212 (150, 288)	<0.001
Highest FiO_2 (%)	50 (40, 80)	50 (40, 90)	50 (40, 80)	<0.001	50 (40, 100)	60 (40, 100)	50 (40, 80)	<0.001
Antibiotic duration (day)	1 (1, 4)	2 (1, 5)	1 (1, 3)	<0.001	0 (0, 2)	1 (0, 4)	0 (0, 2)	<0.001
CRRT duration (day)	0 (0, 0)	0 (0, 0)	0 (0, 0)	<0.001	0 (0, 0)	0 (0, 0)	0 (0, 0)	<0.001
IMV duration (day)	1.5 (0.6, 3.7)	1.9 (0.7, 4.4)	1.2 (0.6, 3.2)	<0.001	2.0 (1.0, 5.0)	3.0 (2.0, 6.0)	2.0 (1.0, 5.0)	<0.001
Urine output (ml/kg/h)	0.6 (0.2, 1.2)	0.5 (0.1, 1.1)	0.7 (0.4, 1.3)	<0.001	0.6 (0.3, 1.1)	0.3 (0.1, 0.8)	0.64 (0.3, 1.1)	<0.001
Vasopressor used 1 day before weaning (*n*, %)	3,882 (77.3)	1,712 (79.3)	2,170 (75.8)	0.004	2,337 (33.0)	511 (54.5)	1,826 (29.7)	<0.001

Values are presented as median and interquartile range (IQR); BMI, body mass index; GCS, Glasgow coma scale; WBC, white blood cell count; PaO_2, arterial oxygen partial pressure; $PaCO_2$, arterial carbon dioxide partial pressure; MAP, mean arterial pressure; SPO_2, pulse oxygen saturation; PEEP, positive end expiratory pressure; OI, oxygenation index; FIO_2, fraction inspired oxygen concentration; CRRT, continuous renal replacement therapy; IMV, invasive mechanical ventilation.

TABLE 2 | The predictive performance of models in the internal and external validation sets.

Models	Internal validation set						External validation set					
	AUROC	Cut-off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	AUROC	Cut-off	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
XGBoost	0.80 (0.77–0.82)	0.57	0.75 (0.71–0.78)	0.71 (0.66–0.75)	0.81 (0.79–0.84)	0.62 (0.58–0.65)	0.86 (0.85–0.87)	0.51	0.96 (0.95–0.96)	0.36 (0.34–0.38)	0.79 (0.79–0.80)	0.77 (0.74–0.79)
The simplified model	0.75 (0.72–0.77)	0.58	0.74 (0.70–0.78)	0.61 (0.56–0.65)	0.68 (0.65–0.70)	0.68 (0.64–0.72)	0.78 (0.77–0.79)	0.54	0.93 (0.93–0.94)	0.32 (0.30–0.34)	0.80 (0.79–0.80)	0.63 (0.60–0.65)
MLP	0.67 (0.64–0.74)	1	0.84 (0.81–0.87)	0.50 (0.46–0.55)	0.84 (0.81–0.86)	0.50 (0.47–0.54)	0.71 (0.69–0.72)	1	0.93 (0.92–0.94)	0.33 (0.30–0.35)	0.81 (0.80–0.81)	0.61 (0.58–0.63)
RF	0.69 (0.66–0.72)	1	0.73 (0.69–0.77)	0.66 (0.61–0.70)	0.76 (0.74–0.79)	0.62 (0.58–0.65)	0.67 (0.66–0.68)	1	0.93 (0.93–0.94)	0.23 (0.21–0.25)	0.64 (0.63–0.64)	0.71 (0.68–0.74)
SVM	0.61 (0.58–0.64)	1	0.64 (0.61–0.67)	0.85 (0.77–0.90)	0.97 (0.95–0.98)	0.26 (0.24–0.29)	0.63 (0.62–0.65)	1	0.90 (0.89–0.91)	0.59 (0.55–0.64)	0.97 (0.96–0.97)	0.30 (0.28–0.32)
LR	0.74 (0.71–0.77)	0.55	0.72 (0.68–0.75)	0.68 (0.63–0.73)	0.81 (0.78–0.83)	0.57 (0.53–0.59)	0.83 (0.82–0.84)	0.58	0.95 (0.94–0.96)	0.33 (0.31–0.35)	0.77 (0.76–0.77)	0.75 (0.73–0.78)
KNN	0.59 (0.56–0.62)	1	0.65 (0.61–0.68)	0.56 (0.50–0.61)	0.74 (0.72–0.77)	0.44 (0.41–0.48)	0.59 (0.58–0.61)	1	0.89 (0.89–0.91)	0.21 (0.19–0.23)	0.75 (0.74–0.75)	0.45 (0.42–0.48)

XGBoost, eXtremely gradient boosting; MLP, multilayer perceptron; RF, random forest; SVM, support vector machine; LR, logistic regression; KNN, KNearest neighbor; AUROC, area under the receiver operating characteristic curves; PPV, positive predictive value; NPV, negative predictive value.

model; the performance of the simplified model is shown in **Figure 4** and **Table 2**. As shown, although the predictive performance of the simplified model decreased slightly, the model was greatly simplified. Subsequently, a web-based tool was developed for clinicians to use a simplified model. The tool can be accessed at http://49.235.211.121/frontdoc/sepweaning.html. After submitting the required data, the probability of weaning was calculated based on the simplified model.

When patients with sepsis are ready to be weaned from invasive mechanical ventilation, clinicians can make a quick decision using the SHAP dependency plots. For scenarios that require precise calculation, the web tool can provide an accurate probability of successful weaning based on the input indicators.

DISCUSSION

This retrospective analysis included two large study cohorts from MIMIC-IV and eICU-CRD. First, we compared mortality and length of ICU stay from WS and WF in patients with sepsis using an effective balance method. Unfortunately, we found significant differences in mortality between patients with WS and WF. Therefore, the XGBoost model and the other five models were applied and evaluated sequentially to identify the beneficial factors associated with WS of patients with sepsis in the ICU. To our knowledge, this is the first study to predict weaning in patients with sepsis based on extensive public data. The difference from previous studies is that we developed an integrated machine learning model with high performance. In addition, we fully evaluated our model using another equally large public dataset. Finally, we explained the main variables and described the effects of their changing trends on weaning.

The matched results showed that there was a significant difference in the length of stay in the ICU between the WS and WF groups (**Figure 1B**, $p = 0.002$). This trend was not confirmed in the external validation set (**Supplementary Figure 4C**). Nevertheless, successful weaning of patients with sepsis significantly reduced mortality (**Figure 1A**, $p < 0.001$), and the matched result from the external validation set showed a similar trend (**Supplementary Figure 4B**). These results were comparable to those of previous studies (18). Therefore, successful weaning is a key factor in improving survival and potentially shortening ICU stay.

In both the internal validation set and the external validation set, our model showed excellent reliability and prospective generalization ability, and the prediction performance was significantly better than that of the traditional method (**Figure 4C**). In the model, the contribution to the prediction of WS varied with the variables. Previous studies have shown that the duration of IMV plays an important role in weaning (19–21). However, because IMV duration is the strongest variable in our model, the contribution of IMV duration to these models is inconsistent. We only included patients with sepsis admitted to the ICU in our study. Therefore, population diversity could lead to this difference. Moreover, it was observed that the best IMV duration was maintained for about 1 day according to the LOWESS

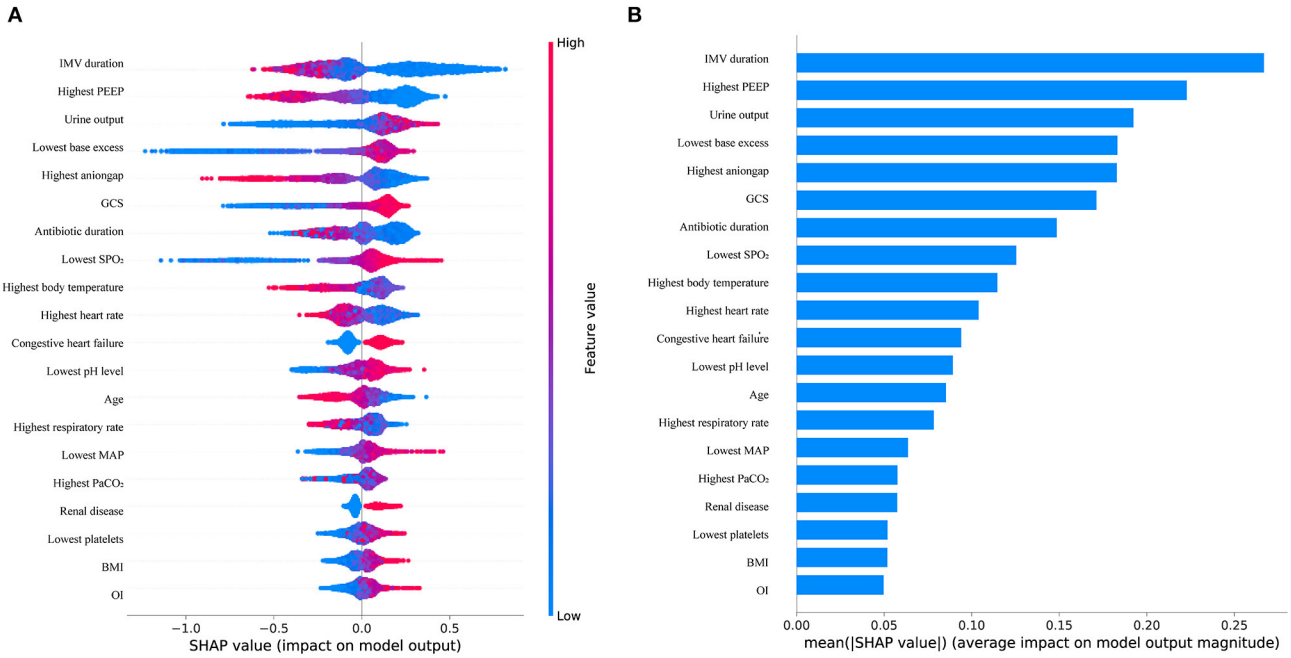


FIGURE 3 | (A) Distribution of the impacts of each variable on the output of the XGBoost model estimated using the SHAP values. **(B)** Ranking of variables importance.

curve (**Figure 5A**). Other variables with prospective cut-off values were antibiotic duration (**Supplementary Figure 2C**), highest body temperature (**Supplementary Figure 2E**), highest heart rate (**Supplementary Figure 2F**), highest respiratory rate (**Supplementary Figure 3A**), and highest $PaCO_2$ (**Supplementary Figure 3C**). Although these variables have also been mentioned in previous studies (19, 21, 22), the trends and specific contributions of these variables have not been clarified.

In our study, age showed a clear downward trend of successful weaning ability with increasing value, especially in patients aged >75 years (**Supplementary Figure 2H**). This trend is supported by increasing evidence (19, 23). Interestingly, BMI and congestive heart failure as low-contributing variables were proportional to WS probability in our study. This finding is supported by previous studies (24, 25). However, there is also an opposite conclusion (19). In a study on viral infections, excessive BMI may lead to worse clinical outcomes (26). However, in another study, even a reasonably high BMI can help patients improve their disease (27). In summary, these differences may be due to the diversity of the population and the diversity of diseases.

PEEP, urine output, and SPO_2, as the most commonly measured indices, play a key role in predicting WS. In this study, a low PEEP strategy was found to be more beneficial for patients with sepsis weaning from ventilation. Although the low-level PEEP strategy did not have a significant effect on improving ventilator weaning in a cohort study (28), it still has the potential to promote ventilator weaning (29). There is increasing evidence that high levels of PEEP are a risk factor in reducing the likelihood of WS (19, 30). As we have considered, high levels of PEEP lead to lung congestion and increase the respiratory burden of patients (31), which may explain why high levels of PEEP play a positive role in the weaning process (**Figure 5B**). Urinary output is the primary means for humans to maintain fluid balance, and excellent fluid management strategies could significantly improve patient survival (32). Polyuria has been shown to have no negative impact on weaning (33), but negative fluid balance significantly impedes weaning success (34, 35). These findings are consistent with those of the present study (**Figure 5C**). The pH was an interesting finding in our study. As the most important indicator of acid-base balance, an abnormally high or low pH had a detrimental effect on weaning in our study. Nevertheless, low pH was more likely to cause weaning failure (**Supplementary Figure 2G**). This finding is consistent with previous studies, as low pH of extracellular fluid compromises the immune function of the body in patients with sepsis (36). BE and $PaCO_2$ are indicators closely related to pH. Early studies have shown that removing $PaCO_2$ from the body effectively improves the success rate of weaning (37, 38). The relationship between BE and weaning has not yet been studied. In any case, extracellular acid-base balance is susceptible to the interaction of several variables. For the three variables of BE, $PaCO_2$, and pH, we used an interactive effects plot to explain the relationship between the three variables. As can be seen in **Supplementary Figure 3**, the tendency of these three variables to predict weaning tended to be consistent, which also shows the rationality of using this model.

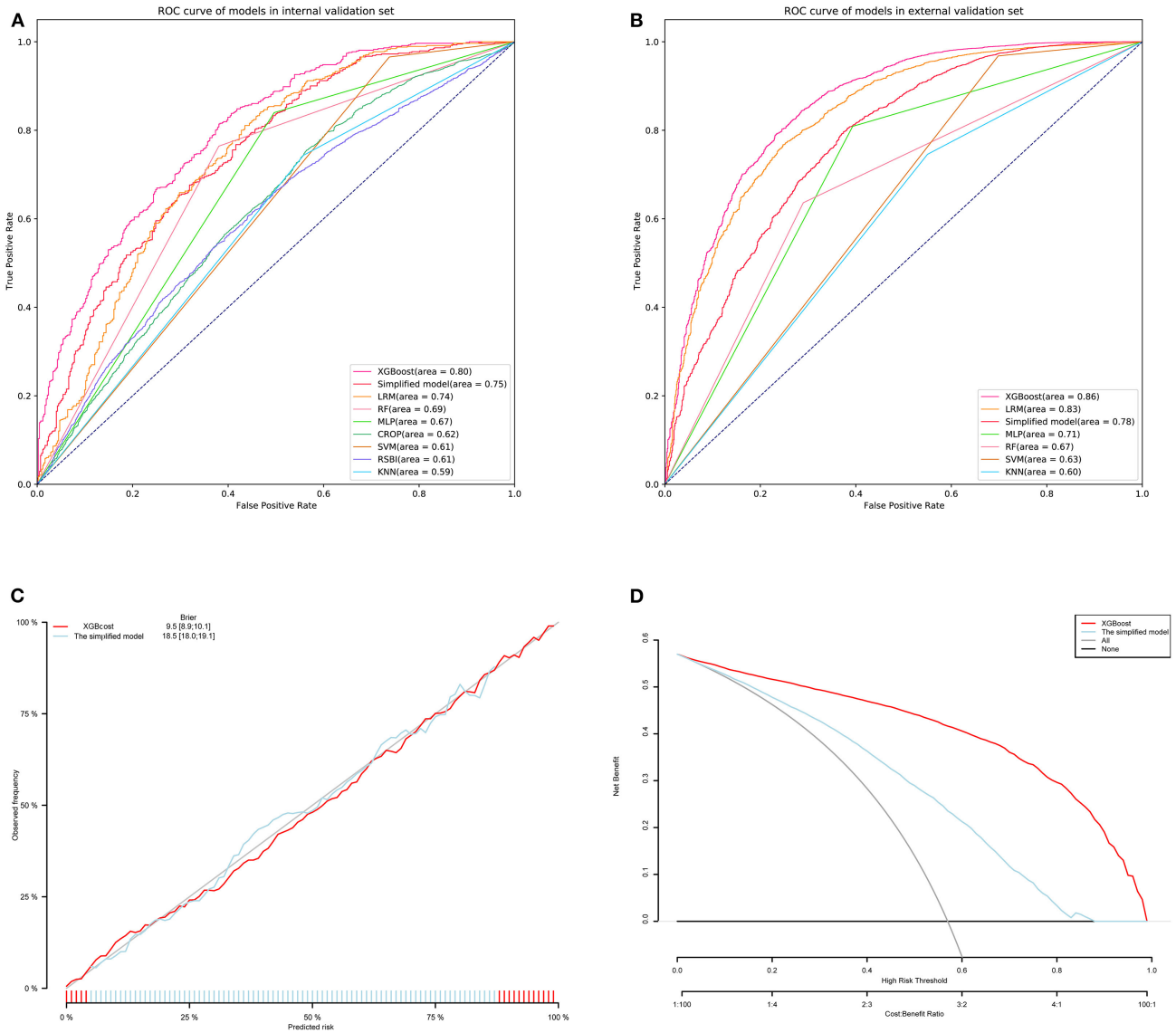


FIGURE 4 | Receiver operating characteristic curves (ROCs) of the XGBoost, LRM, RF, MLP, SVM, KNN, and simplified model. **(A)** Internal validation set. **(B)** External validation set. **(C)** Decision curve analysis of the XGBoost and simplified model. **(D)** Calibration curve of the XGBoost and simplified model. XGBoost, eXtremely gradient boosting; KNN, KNearest neighbor; MLP, multi-layer perceptron; RF, random forest; SVM, support vector machine; LRM, logistic regression; RSBI, rapid shallow breathing Index; CROP, compliance, oxygenation, respiratory rate, pressure index.

In clinical practice, weaning should be a medical behavior that needs careful consideration. Inappropriate weaning may lead to worsening of disease, higher mortality, longer hospital stays, and higher hospital costs (7, 39, 40). Therefore, a simple and effective prediction method is needed. Compared with the traditional prediction model, our model has better prediction performance (11, 41). Although some new weaning models have been proposed in recent years (19, 20, 42, 43), the performance of the model varies according to the target population. In patients with sepsis, the developed model was able to predict weaning well, as reflected by a high AUROC value of 0.80 and 0.86 in the internal and external validation sets, respectively. Interestingly, the performance of the model was better in patients with pulmonary infections than in the validation sets (**Supplementary Figure 5**). Obviously, the performance of our model was better in the external validation set. Apart from the good generalization ability of our model, we believe that the different data sources could explain this phenomenon, which also occurs in other studies (44). Finally, we simplified our model and developed a web-based tool that allows convenient use of the model.

There are still potential limitations in our research. First, because of the limitation of the database, other variables that could have predictive value, such as lactate and central

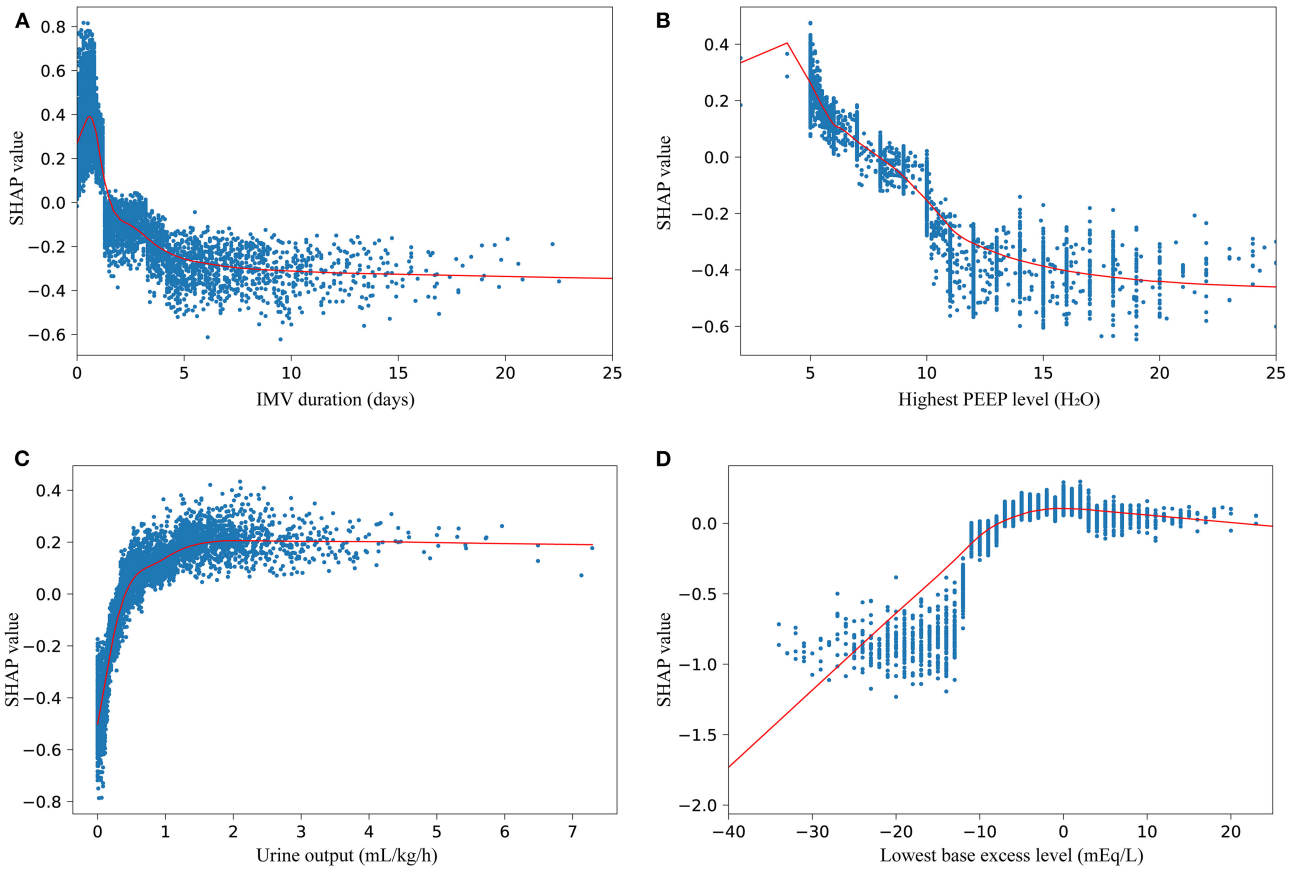


FIGURE 5 | SHAP dependency plots of IMV duration **(A)**, highest PEEP level **(B)**, urine output **(C)**, and lowest base excess level **(D)**.

venous pressure, with excessively high error rates, were not included in the model. Considering the availability of comprehensive clinical indicators, such as SOFA and SAPS, were not included in the model, although these indicators could improve the predictive performance of the model (19, 45, 46). Second, although our model was validated using data from multiple data sources, we still need additional data sources to further demonstrate the generalizability of the model. Third, in our study, the positive predictive value (PPV) and negative predictive value (NPV) were 0.81, 0.62, 0.79, and 0.77 in the internal and external validation sets, respectively, which means that the model still has some degree of false-positive and false-negative rates. More valuable variables and dynamic prediction models could improve the performance.

In conclusion, weaning success is independent of short-term mortality in patients with sepsis. We developed a prospective model for weaning from invasive mechanical ventilation using the XGBoost algorithm. This model included 35 conventional clinical variables and proved to be more interpretable and predictive. In addition, the model was simplified, and a web-based tool was developed for better clinical application.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://physionet.org/about/database/.

AUTHOR CONTRIBUTIONS

MY and WL carried out the concepts, design, data acquisition, and manuscript preparation. GT, ZL, and YZ carried out literature search and manuscript preparation. WX, JZ, YL, and TH performed manuscript review, including revision of key technical content and English expression. All authors have read and approved the content of the manuscript.

FUNDING

This work is supported by National Natural Science Foundation of China (Grant No. 82072134).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.814566/full#supplementary-material

REFERENCES

1. Zilberberg MD, Nathanson BH, Ways J, Shorr AF. Characteristics, hospital course, and outcomes of patients requiring prolonged acute versus short-term mechanical ventilation in the United States, 2014-2018. *Crit Care Med.* (2020) 48:1587–94. doi: 10.1097/CCM.0000000000004525
2. Amoateng-Adjepong Y, Jacob BK, Ahmad M, Manthous CA. The effect of sepsis on breathing pattern and weaning outcomes in patients recovering from respiratory failure. *Chest.* (1997) 112:472–7. doi: 10.1378/chest.112.2.472
3. Nin N, Lorente JA, Fernández-Segoviano P, De Paula M, Ferruelo A, Esteban A. High-tidal volume ventilation aggravates sepsis-induced multiorgan dysfunction in a dexamethasone-inhibitable manner. *Shock.* (2009) 31:429–34. doi: 10.1097/SHK.0b013e318188b720
4. Wellman TJ, Winkler T, Costa EL, Musch G, Harris RS, Zheng H, et al. Effect of local tidal lung strain on inflammation in normal and lipopolysaccharide-exposed sheep*. *Crit Care Med.* (2014) 42:e491–500. doi: 10.1097/CCM.0000000000000346
5. Motta-Ribeiro GC, Hashimoto S, Winkler T, Baron RM, Grogg K, Paula LFSC, et al. Deterioration of Regional Lung Strain and Inflammation during Early Lung Injury. *Am J Respir Crit Care Med.* (2018) 198:891–902. doi: 10.1164/rccm.201710-2038OC
6. Fernandez-Zamora MD, Gordillo-Brenes A, Banderas-Bravo E, Arboleda-Sánchez JA, Hinojosa-Pérez R, Aguilar-Alonso E, et al. Prolonged mechanical ventilation as a predictor of mortality after cardiac surgery. *Respir Care.* (2018) 63:550–7. doi: 10.4187/respcare.04915
7. Frutos-Vivar F, Esteban A, Apezteguia C, González M, Arabi Y, Restrepo MI, et al. Outcome of reintubated patients after scheduled extubation. *J Crit Care.* (2011) 26:502–9. doi: 10.1016/j.jcrc.2010.12.015
8. Saiphoklang N, Auttajaroon J. Incidence and outcome of weaning from mechanical ventilation in medical wards at Thammasat University Hospital. *PLoS ONE.* (2018) 13:e0205106. doi: 10.1371/journal.pone.0205106
9. Frutos-Vivar F, Ferguson ND, Esteban A, Epstein SK, Arabi Y, Apezteguía C, et al. Risk factors for extubation failure in patients following a successful spontaneous breathing trial. *Chest.* (2006) 130:1664–71. doi: 10.1378/chest.130.6.1664
10. Sassoon CS, Mahutte CK. Airway occlusion pressure and breathing pattern as predictors of weaning outcome. *Am Rev Respir Dis.* (1993) 148:860–6. doi: 10.1164/ajrccm/148.4_Pt_1.860
11. Yang KL, Tobin MJ. A prospective study of indexes predicting the outcome of trials of weaning from mechanical ventilation. *N Engl J Med.* (1991) 324:1445–50. doi: 10.1056/NEJM199105233242101
12. Hilbert G, Gruson D, Portel L, Vargas F, Gbikpi-Benissan G, Cardinaud JP. Airway occlusion pressure at 01 s (P01) after extubation: an early indicator of postextubation hypercapnic respiratory insufficiency. *Intensive Care Med.* (1998) 24:1277–82. doi: 10.1007/s001340050762
13. Jubran A, Tobin MJ. Pathophysiologic basis of acute respiratory distress in patients who fail a trial of weaning from mechanical ventilation. *Am J Respir Crit Care Med.* (1997) 155:906–15. doi: 10.1164/ajrccm.155.3.9117025
14. Moschietto S, Doyen D, Grech L, Dellamonica J, Hyvernat H, Bernardin G. Transthoracic Echocardiography with Doppler Tissue Imaging predicts weaning failure from mechanical ventilation: evolution of the left ventricle relaxation rate during a spontaneous breathing trial is the key factor in weaning outcome. *Crit Care.* (2012) 16:R81. doi: 10.1186/cc11339
15. Shankar-Hari M, Phillips GS, Levy ML, Seymour CW, Liu VX, Deutschman CS, et al. Developing a new definition and assessing new clinical criteria for septic shock: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA.* (2016) 315:775–87. doi: 10.1001/jama.2016.0289
16. Schönhofer B, Geiseler J, Dellweg D, Fuchs H, Moerer O, Weber-Carstens S, et al. Prolonged weaning: S2k guideline published by the German Respiratory Society. *Respiration.* (2020) 2020:1-102. doi: 10.1159/000510085
17. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model.* (1986) 7:1393–512. doi: 10.1016/0270-0255(86)90088-6
18. Yang CH, Hsiao JL, Wu MF, Lu MH, Chang HM, Ko WS, et al. The declined levels of inflammatory cytokines related with weaning rate during period of septic patients using ventilators. *Clin Respir J.* (2018) 12:772–8. doi: 10.1111/crj.12593
19. Zhao QY, Wang H, Luo JC, Luo MH, Liu LP, Yu SJ, et al. Development and validation of a machine-learning model for prediction of extubation failure in intensive care units. *Front Med.* (2021) 8:676343. doi: 10.3389/fmed.2021.676343
20. Goel N, Chakraborty M, Watkins WJ, Banerjee S. Predicting extubation outcomes-a model incorporating heart rate characteristics index. *J Pediatr.* (2018) 195:53–58.e1. doi: 10.1016/j.jpeds.2017.11.037
21. Thille AW, Boissier F, Ben Ghezala H, Razazi K, Mekontso-Dessap A, Brun-Buisson C. Risk factors for and prediction by caregivers of extubation failure in ICU patients: a prospective study. *Crit Care Med.* (2015) 43:613–20. doi: 10.1097/CCM.0000000000000748
22. Minozzi S, Pifferi S, Brazzi L, Pecoraro V, Montrucchio G, D'Amico R. Topical antibiotic prophylaxis to reduce respiratory tract infections and mortality in adults receiving mechanical ventilation. *Cochrane Database Syst Rev.* (2021) 1:CD000022. doi: 10.1002/14651858.CD000022.pub4
23. Warnke C, Heine A, Müller-Heinrich A, Knaak C, Friesecke S, Obst A, et al. Predictors of survival after prolonged weaning from mechanical ventilation. *J Crit Care.* (2020) 60:212–7. doi: 10.1016/j.jcrc.2020.08.010
24. Keng LT, Liang SK, Tseng CP, Wen YF, Tsou PH, Chang CH, et al. Functional status after pulmonary rehabilitation as a predictor of weaning success and survival in patients requiring prolonged mechanical ventilation. *Front Med.* (2021) 8:675103. doi: 10.3389/fmed.2021.675103
25. Nguyen Q, Coghlan K, Hong Y, Nagendran J, MacArthur R, Lam W. Factors associated with early extubation after cardiac surgery: a retrospective single-center experience. *J Cardiothorac Vasc Anesth.* (2021) 35:1964–70. doi: 10.1053/j.jvca.2020.11.051
26. Chand S, Kapoor S, Orsi D, Fazzari M, Tanner T, Umeh G, et al. COVID-19-associated critical illness-report of the first 300 patients admitted to intensive care units at a New York City Medical Center. *J Intensive Care Med.* (2020) 35:963–70. doi: 10.1177/0885066620946692
27. Sakr Y, Alhussami I, Nanchal R, Wunderink RG, Pellis T, Wittebole X, et al. Being overweight is associated with greater survival in ICU patients: results from the intensive care over nations audit. *Crit Care Med.* (2015) 43:2623–32. doi: 10.1097/CCM.0000000000001310
28. Zhao H, Su L, Ding X, Chen H, Zhang H, Wang J, et al. The risk factors for weaning failure of mechanically ventilated patients with COVID-19: a retrospective study in national medical team work. *Front Med.* (2021) 8:678157. doi: 10.3389/fmed.2021.678157
29. Writing Committee and Steering Committee for the RELAx Collaborative Group, Algera AG, Pisani L, Serpa Neto A, den Boer SS, Bosch FFH, et al. Effect of a lower vs higher positive end-expiratory pressure strategy on ventilator-free days in ICU patients without ARDS: a randomized clinical trial. *JAMA.* (2020) 324:2509–20. doi: 10.1001/jama.2020.23517
30. Ionescu F, Zimmer MS, Petrescu I, Castillo E, Bozyk P, Abbas A, et al. Extubation failure in critically ill COVID-19 patients: risk factors and impact on in-hospital mortality. *J Intensive Care Med.* (2021) 36:1018–24. doi: 10.1177/08850666211020281
31. Retamal J, Bugedo G, Larsson A, Bruhn A. High PEEP levels are associated with overdistension and tidal recruitment/derecruitment in ARDS patients. *Acta Anaesthesiol Scand.* (2015) 59:1161–9. doi: 10.1111/aas.12563
32. Zhang Z, Zheng B, Liu N. Individualized fluid administration for critically ill patients with sepsis with an interpretable dynamic treatment regimen model. *Sci Rep.* (2020) 10:17874. doi: 10.1038/s41598-020-74906-z
33. Aliyali M, Sharifpour A, Abedi S, Spahbodi F, Namarian N, Zarea A, et al. The ability of polyuria in prediction of weaning outcome in critically ill mechanically ventilated patients. *Tanaffos.* (2019) 18:74–8.
34. Maezawa S, Kudo D, Miyagawa N, Yamanouchi S, Kushimoto S. Association of body weight change and fluid balance with extubation failure in intensive care unit patients: a single-center observational study. *J Intensive Care Med.* (2021) 36:175–81. doi: 10.1177/0885066619887694
35. Lin MY, Li CC, Lin PH, Wang JL, Chan MC, Wu CL, et al. Explainable machine learning to predict successful weaning among patients requiring prolonged mechanical ventilation: a retrospective cohort study in Central Taiwan. *Front Med.* (2021) 8:663739. doi: 10.3389/fmed.2021.663739
36. Lardner A. The effects of extracellular pH on immune function. *J Leukoc Biol.* (2001) 69:522-30. doi: 10.1189/jlb.69.4.522

37. Abrams DC, Brenner K, Burkart KM, Agerstrand CL, Thomashow BM, Bacchetta M, et al. Pilot study of extracorporeal carbon dioxide removal to facilitate extubation and ambulation in exacerbations of chronic obstructive pulmonary disease. *Ann Am Thorac Soc.* (2013) 10:307–14. doi: 10.1513/AnnalsATS.201301-021OC
38. Braune S, Sieweke A, Brettner F, Staudinger T, Joannidis M, Verbrugge S, et al. The feasibility and safety of extracorporeal carbon dioxide removal to avoid intubation in patients with COPD unresponsive to noninvasive ventilation for acute hypercapnic respiratory failure (ECLAIR study): multicentre case-control study. *Intensive Care Med.* (2016) 42:1437–44. doi: 10.1007/s00134-016-4452-y
39. Thille AW, Harrois A, Schortgen F, Brun-Buisson C, Brochard L. Outcomes of extubation failure in medical intensive care unit patients. *Crit Care Med.* (2011) 39:2612–8. doi: 10.1097/CCM.0b013e3182282a5a
40. Perren A, Previsdomini M, Llamas M, Cerutti B, Györik S, Merlani G, et al. Patients' prediction of extubation success. *Intensive Care Med.* (2010) 36:2045–52. doi: 10.1007/s00134-010-1984-4
41. Hendra KP, Bonis PA, Joyce-Brady M. Development and prospective validation of a model for predicting weaning in chronic ventilator dependent patients. *BMC Pulm Med.* (2003) 3:3. doi: 10.1186/1471-2466-3-3
42. Vassilakopoulos T, Routsi C, Sotiropoulou C, Bitsakou C, Stanopoulos I, Roussos C, et al. The combination of the load/force balance and the frequency/tidal volume can predict weaning outcome. *Intensive Care Med.* (2006) 32:684–91. doi: 10.1007/s00134-006-0104-y
43. Jia Y, Kaul C, Lawton T, Murray-Smith R, Habli I. Prediction of weaning from mechanical ventilation using convolutional neural networks. *Artif Intell Med.* (2021) 117:102087. doi: 10.1016/j.artmed.2021.102087
44. Chen Y, Jiang S, Lu Z, Xue D, Xia L, Lu J, et al. Development and verification of a nomogram for prediction of recurrence-free survival in clear cell renal cell carcinoma. *J Cell Mol Med.* (2020) 24:1245–55. doi: 10.1111/jcmm.14748
45. Brix N, Sellmer A, Jensen MS, Pedersen LV, Henriksen TB. Predictors for an unsuccessful INtubation-SURfactant-Extubation procedure: a cohort study. *BMC Pediatr.* (2014) 14:155. doi: 10.1186/1471-2431-14-155
46. Jaber S, Quintard H, Cinotti R, Asehnoune K, Arnal JM, Guitton C, et al. Risk factors and outcomes for airway failure versus non-airway failure in the intensive care unit: a multicenter observational study of 1514 extubation procedures. *Crit Care.* (2018) 22:236. doi: 10.1186/s13054-018-2150-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Medicine

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fmed.2021.813640

Emerging Trends and Hot Spots of Electrical Impedance Tomography Applications in Clinical Lung Monitoring

Zhe Li[1†], Shaojie Qin[1†], Chen Chen[1], Shuya Mei[1], Yulong Yao[1], Zhanqi Zhao[2,3], Wen Li[1], Yuxiao Deng[1*] and Yuan Gao[1*]*

[1] Department of Critical Care Medicine, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [2] Department of Biomedical Engineering, Fourth Military Medical University, Xi'an, China, [3] Institute of Technical Medicine, Furtwangen University, Villingen-Schwenningen, Germany

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Songqiao Liu,
Southeast University, China
Ming Zhong,
Fudan University, China

****Correspondence:***
Wen Li
ciwenzi@sina.com
Yuxiao Deng
dengyuxiao@renji.com
Yuan Gao
rj_gaoyuan@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *12 November 2021*
Accepted: *28 December 2021*
Published: *31 January 2022*

Citation:
Li Z, Qin S, Chen C, Mei S, Yao Y, Zhao Z, Li W, Deng Y and Gao Y (2022) Emerging Trends and Hot Spots of Electrical Impedance Tomography Applications in Clinical Lung Monitoring. Front. Med. 8:813640. doi: 10.3389/fmed.2021.813640

Objective: This study explores the emerging trends and hot topics concerning applications on electrical impedance tomography (EIT) in clinical lung monitoring.

Methods: Publications on EIT applications in clinical lung monitoring in 2001–2021 were extracted from the Web of Science Core Collection (WoSCC). The search strategy was "electrical impedance tomography" and "lung." CiteSpace, a VOS viewer was used to study the citation characteristics, cooperation, and keyword co-occurrence. Moreover, co-cited reference clustering, structural variation analysis (SVA), and future research trends were presented.

Results: Six hundred and thirty-six publications were included for the final analysis. The global annual publications on clinical lung monitoring gradually increased in the last two decades. Germany contributes 32.2% of total global publications. University Medical Center Schleswig-Holstein (84 publications, cited frequency 2,205), *Physiological Measurement* (105 publications, cited frequency 2,056), and Inéz Frerichs (116 articles, cited frequency 3,609) were the institution, journal, and author with the largest number of article citations in the research field. "Electrical impedance tomography" (occurrences, 304), "mechanical ventilation" (occurrences, 99), and "acute respiratory distress syndrome" (occurrences, 67) were the top most three frequent keywords, "noninvasive monitoring" (Avg, pub, year: 2008.17), and "extracorporeal membrane oxygenation" (Avg, pub, year: 2019.60) were the earliest and latest keywords. The keywords "electrical impedance tomography" (strength 7.88) and co-cited reference "Frerichs I, 2017, THORAX" (strength 47.45) had the highest burst value. "Driving pressure," "respiratory failure," and "titration" are the three keywords still maintaining a high brush value until now. The largest and smallest cluster of the co-cited references are "obstructive lung diseases" (#0, size: 97) and "lung perfusion" (#20, size: 5). Co-cited reference "Frerichs I, 2017, THORAX" (modularity change rate: 98.49) has the highest structural variability. Categories with most and least interdisciplinary crossing are "ENGINEERING" and "CRITICAL CARE MEDICINE."

Conclusions: EIT is a valuable technology for clinical lung monitoring, gradually converting from imaging techniques to the clinic. Research hot spots may continue monitoring techniques, the ventilation distribution of acute respiratory distress syndrome (ARDS), and respiratory therapy strategies. More diversified lung function monitoring studies, such as lung perfusion and interdisciplinary crossing, are potentially emerging research trends.

Keywords: bibliometric analysis, EIT, lung, ARDS, hotspot

INTRODUCTION

Electrical impedance tomography (EIT) is a non-invasive, radiation-free functional imaging technology invented over three decades ago, with the real-time application of monitoring global and regional lung function and ventilation distribution at the bedside (1). Recently, in the development of evidence-based medicine, a growing number of studies have confirmed that EIT is a useful tool in optimizing individual ventilator parameters, improving gas exchange, increasing oxygen levels, and decreasing ventilator-induced lung injury in respiratory failure (2–4). Clinical trials have reported that patients with acute respiratory distress syndrome (ARDS) could benefit from EIT-guided respiratory therapy (5, 6). With the clinical experience and the progress of clinical studies, topics regarding EIT clinical applications have been presented in different professional journals and scientific conferences. Much more attention has been given to the clinical applications of EIT. An intuitive overview and explicit research trends of clinical applications of EIT are beneficial for researchers to improve knowledge uptake, identify scientific advances, hot spots, research trends, and cooperative relationships as well as promote interdisciplinary cooperation. However, the studies that show the emerging trends and hot spots of publications in the field have not been reported.

A bibliometric analysis analyzes research publications based on big data and artificial intelligence (AI) (7). Research publications play an essential role in transmitting the process information of scientific development in a certain research field. AI has an innate advantage in dealing with huge amounts of data, e.g., enormous publications, and is more convenient in interpreting different quantitative rules in the network world. Therefore, a bibliometric analysis is used to evaluate contributions to a research field, including those by countries, institutions, authors, and journals (8). With the continuous improvement in the performance and effect of machine learning, AI has promoted the upgrading of bibliometric technology going deeper into investigation hot spots and research trends (9).

To explore the current status, emerging trends, and hot spots of EIT in clinical applications, a bibliometric analysis was conducted on the topic for the past two decades. This study helps researchers identify the most significant and impactful articles that highlight the characteristics of EIT application in clinical lung monitoring, and provide valuable insights into the most noteworthy research landscape, and forecast future work.

METHODS

Data Sources and Search Strategies

The Web of Science (WoS) database, which has a rich collection of scientific literature, is commonly chosen for bibliometric analysis. In this study, all data were retrieved from the Web of Science Core Collection (WoSCC). The search strategy was TS = ("electrical impedance tomography") AND ("lung"), Time window: January 1, 2001, to May 29, 2021, Publication type: "Article," "Review," and "Letter," Language: English. All investigators collected the literature on May 29, 2021, to avoid database update bias.

Data Collection

Two investigators (QSJ and LZ) independently extracted all data, including publications, author, title, abstract, keywords, source, language, citation. Publications less relevant to clinical applications are defined as studies and reviews based on image processing, algorithm, equipment assembly not limited to the electrode, and belt optimization or exploring the monitoring of other biological directions. The data were saved in a text and a UTF-8 format from the WoS core collection and saved for further software analysis.

Bibliometric Analysis

WoS Core Database Output Analysis

The intrinsic functions of the WoS core database and Microsoft Excel (version Microsoft 365) were used to describe the features of the publications, including the total number of literature, annual, national, institutional, individual article counts, research field distributions, and top-cited literature.

Network Analysis

Our data were imported into the bibliometric analysis software VOS (VOS viewer 1.6.16, Leiden University, Leiden, the Netherlands) or CiteSpace 5.7R5 (Chen Meichao, Drexel University) using a UTF-8 or a text format for network analysis. VOS finished top citations, coauthorship, co-occurrence analysis of authors, institutions, countries, keywords, and other factors. The numbers of documents and citations and the strength of the links were recorded and visually rendered with corresponding symbol color and line weight. The reference cluster analysis, citation bursts, and timeline view analysis, structural variation analysis (SVA), category co-occurrence analysis, etc., were finished using CiteSpace. Cluster type, size, strength timeline, and landmark literature were analyzed and visual rendering with

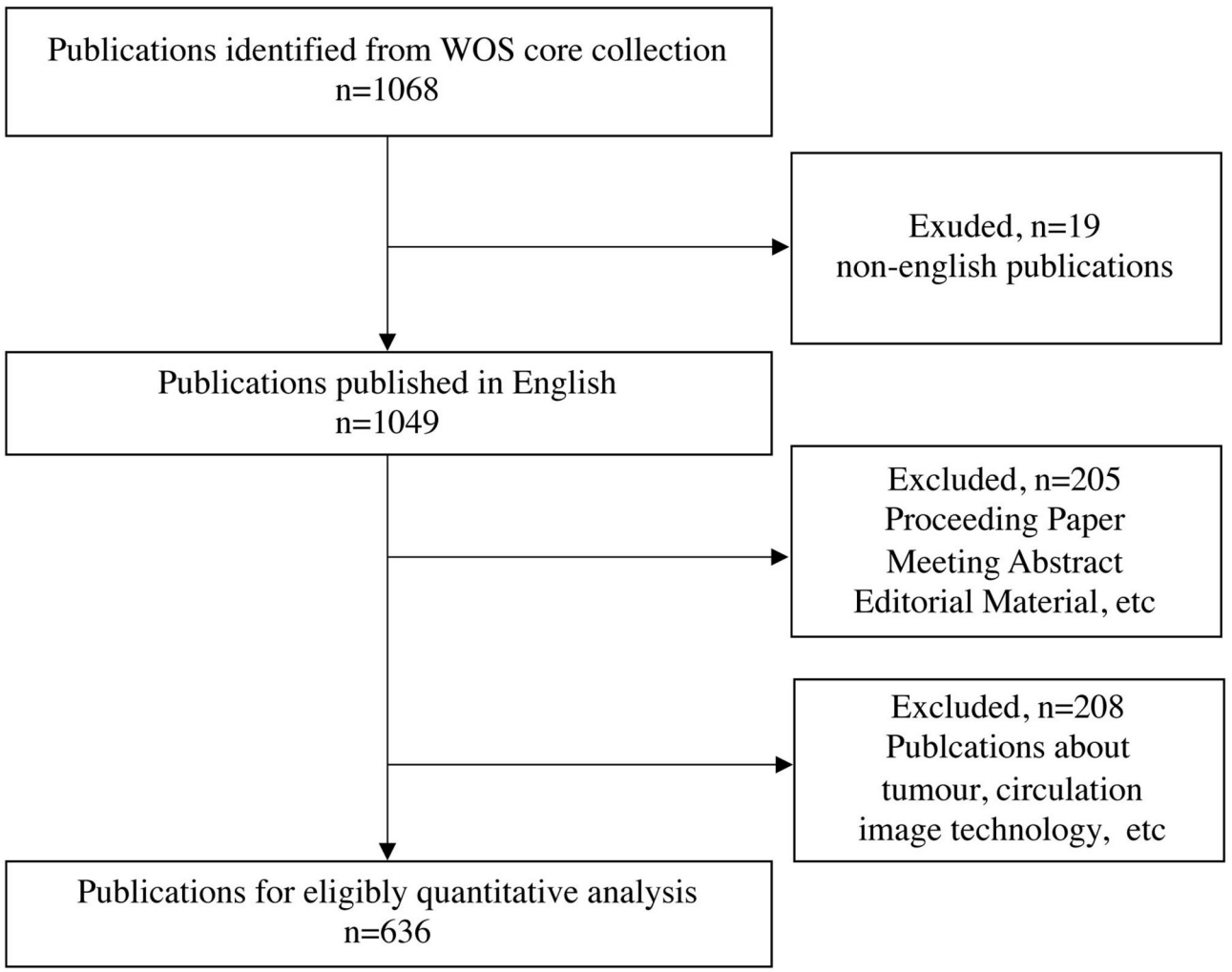


FIGURE 1 | Flow chart of included publications. WoSC, Web of science center.

corresponding image features (10). Keyword timeline view and keyword strongest citation burst analysis were conducted to show the topic's time scale and strength. The SVA was conducted to show landmark literature's contribution to this field. The category analysis was used to evaluate disciplinary cooperation.

RESULTS

Bibliometric Analysis of Publication Output

Following the retrieval strategy, 1,068 publications were identified and further screened. Finally, 636 research articles, reviews, and letters published in English on EIT applications in clinical lung monitoring were analyzed, and 432 publications were excluded (**Figure 1**).

Growth Trend of Publications and Global Cooperation

The literature counts between 2001 and 2020 illustrated the growth trend and global geographic distribution of publications in the field. Following the WoSCC database, 48 countries contributed to publications on EIT related to clinical lung monitoring. The global and top 10 countries in publications are indicated in **Figure 2A**. Overall, the annual EIT global publication number is increasing. Germany contributed the largest number of publications and has the most active cooperation on EIT applications in clinical lung monitoring studies with other countries (**Figure 2B**). Global national cooperation and countries collaborating most with Germany were indicated in **Supplementary Figure 1**.

Analysis of Top Publications and Reference Articles

The ranking of the top 10 institutions, journals, authors of publications, and reference articles cited using EIT clinical lung monitoring publications is indicated in **Table 1**. Universities and their affiliated hospitals are the primary institution

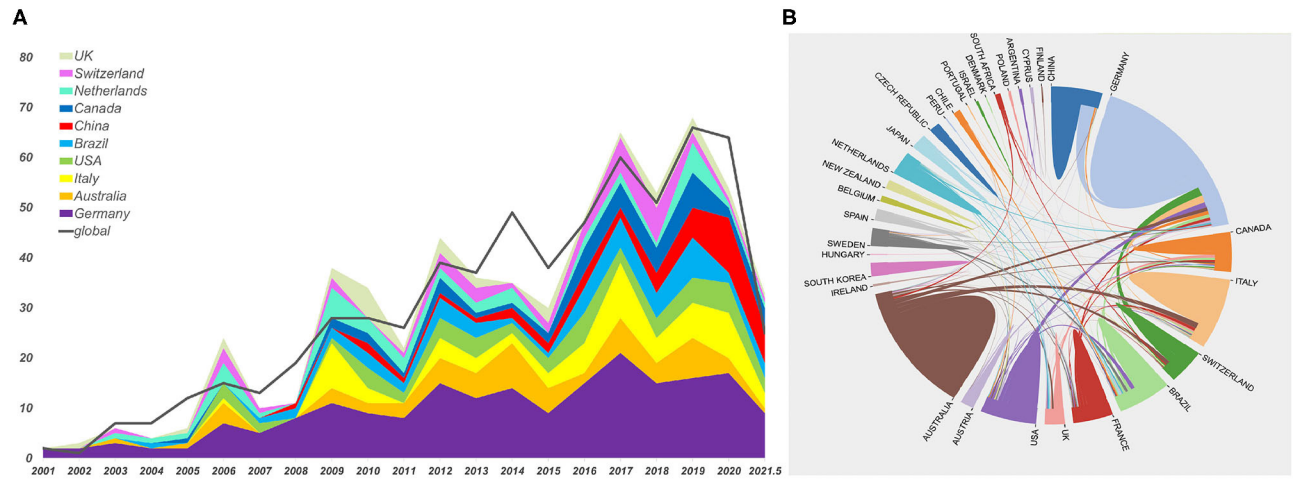


FIGURE 2 | The number of annual publications, growth trends **(A)**, and country coauthorship analysis **(B)** of global and top 10 countries on electrical impedance tomography (EIT) applications in clinical lung monitoring research from 2001 to May 29, 2021. **(A)** The gray line indicated the trend of the annual world publication number. The color indicated different countries. The width of the color plate indicated the annual publication number of each country. **(B)** The color indicated countries; color plate size indicated the number of publications; the thickness of lines indicated linkage strength.

TABLE 1 | The top 10 institutions, journals, authors of publications, and reference articles on electrical impedance tomography (EIT) lung monitoring.

Rank	Institution	Publications	Citations	Journal	Publications	Citations	Author	Publications	Citations	Cited reference	Citations
1	Univ Med Ctr Schleswig Holstein	84	2,205	Physiol Meas	105	2,056	Frerichs, I	116	3,609	Victorino JA, 2004, AM J RESP CRIT CARE (11)##	196
2	Univ São Paulo	49	2,061	Intens Care Med	31	1,723	Zhao, ZQ	43	896	FrerichsI, 2017, THORAX (12)*	147
3	Univ Milan	37	803	Crit Care Med	23	1,115	Adler, A	34	1,113	Frerichs I, 2002, J APPL PHYSIOL (13)#	138
4	Furtwangn Univ	57	1,000	Crit Care	33	784	Leonhardt, S	28	814	Adler A, 2009, PHYSIOL MEAS (14)*	109
5	RWTH Aachen Univ	35	1,044	Curr Opin Crit Care	22	625	Tingay, D	27	628	Zhao ZQ, 2009, INTENS CARE MED (15)##	98
6	Royal Childrens Hospi	32	662	J Appl Physiol	18	586	Weiler, N	25	991	Meier T, 2008, INTENS CARE MED (16)#	95
7	Carleton Univ	38	1,152	Acta Anaesth Scand	25	360	Mauri, T	23	839	Frerichs I, 2000, Physiol Meas (17)*	94
8	Univ Melbourne	31	440	Resp Care	18	287	Amato, M	21	1,297	Hinz j, 2003, CHEST (18)#	92
9	Fourth mil Med Univ	30	134	PLoS ONE	17	162	Schibler, A	20	666	Brower RG, 2000, New Engl J Med (19)##	89
10	Murdoch Childrens Res Inst	29	602	Am J Resp Crit Care	16	1,128	Moeller, K	35	554	Frerichs I, 2006, Am J Resp Crit Care (20)#	86

Cited reference: the references articles of publications on EIT applications in clinical lung monitoring involved in the analysis.

**Review.*

#Animal research.

##Clinical research.

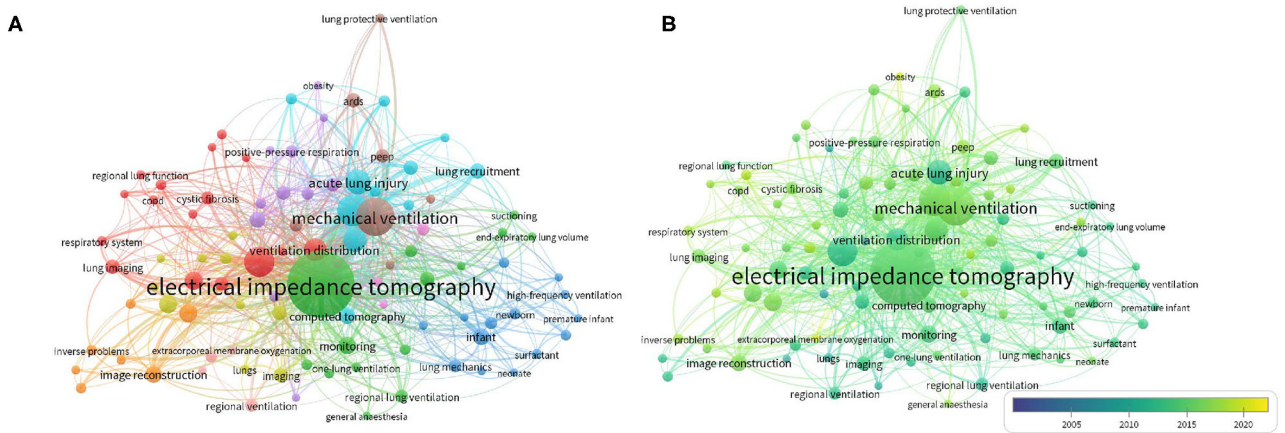


FIGURE 3 | The keyword co-occurrence network and overlay analysis on EIT lung monitoring research from 2001 to May 29, 2021. **(A)** Network visualization of the keyword co-occurrence. **(B)** Overlay visualization of keyword co-occurrence. The color indicated various clusters **(A)** and the average publication year **(B)**. The circle size indicated the number of occurrences, and the thickness of lines indicated the linkage strength. The color indicated the average publication year of keywords. Circle size indicated the number of occurrences; the thickness of lines indicated the strength of the linkage. The distance between circles indicated their relationship.

types contributed to publications in our research field. The department of University Medical Center Schleswig-Holstein (Univ Med Ctr Schleswig-Holstein) (publications: 84 and citing frequency: 2,205) owns most publications, Furtwangen University (Furtwangen Univ) (publications: 57 and citing frequency: 1,000) and University of São Paulo (Univ São Paulo) (publications: 49 and citing frequency: 2,061) ranked the second and third. Fourth Military Medical University (Fourth mil Med Univ) (publications: 30 and citing frequency: 134) ranked ninth and was the only Chinese institution to enter the top ten. Close cooperation can be observed in most productive and cited institutions (**Supplementary Figure 1**). The journal of Physiological Measurement (PHYSIOL MEAS) (105 publications, cited frequency 2,056), Critical Care (CRIT CARE) (33 publications, cited frequency 784), Intensive Care Medicine (INTENS CARE MED) (31 publications, cited frequency 1,723), Acta Anaesthesiologica Scandinavica (ACTA ANAESTH SCAND; 25 publications, cited frequency 360), and Critical Care Medicine (CRIT CARE MED; 23 publications, cited frequency 1,115) had the top five numbers of EIT publications in lung monitoring. Three of them are critical care-specialized journals. Inéz Frerichs (Frerichs, I; 116 publications, cited frequency 3,609) published most publications in the research field, followed by Zhao Zhanqi (Zhao ZQ; 43 publications, cited frequency 896), and Andy Adler (Adler A; 34 publications, cited frequency 1,113). Frerichs owned the largest author cooperative relationships as the most productive author. Furthermore, active cooperation could be found between high publishing and cited authors (**Supplementary Figure 1**). The top 10 most cited reference articles consisted of three reviews, four animal studies, and three clinical studies, "Victorino JA, 2004, AM J RESP CRIT CARE" (cited frequency 196), a clinical study, was a reference article with the most cited frequency (11).

TABLE 2 | Top 10 highly frequent keywords from the included publications on EIT lung monitoring ($n = 100$).

Rank	Keyword	Occurrences	Total link strengths
1	Electrical impedance tomography	304	557
2	Mechanical ventilation	99	259
3	Acute respiratory distress syndrome	67	188
4	Positive end-expiratory pressure	38	116
5	Ventilation distribution	35	89
6	Image reconstruction	19	48
7	Infant	19	56
8	Lung recruitment	16	49
9	Ventilator-induced lung injury	14	39
10	Cystic fibrosis	12	22

Bibliometric Analysis of Keyword Co-occurrence and Topic Trends

Keyword Co-occurrence Network and Overlay Analysis

One thousand two hundred keywords were identified from the included articles. A total of 100 keywords that occurred five or more times were defined as high-frequency keywords and enrolled in a co-occurrence network and overlay analysis. The keywords co-occurred in 10 clusters, 1,578 links, and 3,560 total link strengths (**Figure 3A**). "EIT" (occurrences, 304; total link strengths, 557) takes the top of highly frequency keyword, with a strong co-occurrence to "mechanical ventilation" (occurrences, 99; link strengths with EIT, 46), "acute respiratory distress syndrome" (occurrences, 67; link strengths with EIT, 40), "acute lung injury" (occurrences, 42; link strengths with EIT, 22), and "positive end-expiratory pressure (PEEP)" (occurrences, 38; link strengths with EIT, 19). The next top four high-frequency

TABLE 3 | The top 10 earliest and latest keywords of included publications on EIT lung monitoring ($n = 100$).

Earliest				Latest		
Rank	**Keyword**	**Occurrences**	**Avg, pub, year**	**Keyword**	**Occurrences**	**Avg, pub, year**
1	Non-invasive monitoring	6	2008.17	Extracorporeal membrane oxygenation	5	2019.60
2	Pulmonary edema	5	2008.60	Obesity	5	2019.00
3	Intensive care unit	5	2009.60	Lung perfusion	6	2017.67
4	Regional lung ventilation	11	2011.00	High-flow nasal cannula	6	2017.33
5	Premature infant	5	2011.40	Overdistention	6	2017.33
6	Diagnostic imaging	8	2011.75	Pulmonary function testing	5	2017.20
7	Regularization	6	2011.83	Transpulmonary pressure	6	2017.17
8	Regional ventilation	11	2011.91	peep	13	2017.00
9	Ventilation distribution	35	2012.03	Acute respiratory distress syndrome	5	2017.00
10	Lung impedance	7	2012.14	General anesthesia	5	2017.00

Avg, pub, year: Average co-occurrence year (to the nearest 2 decimal place).

keywords were "mechanical ventilation" (occurrences, 99; total link strength, 259), "acute respiratory distress syndrome" (occurrences, 67; total link strength, 188), "PEEP" (occurrences, 38; total link strength, 116), and "ventilation distribution" (occurrences, 35; total link strength, 89). The top ten highly frequent keywords are ranked in **Table 2**.

The overlay analysis of the keywords representing the topic trends of EIT applications in clinical lung monitoring. The color distribution of keywords in the overlay visual map showed different periods (**Figure 3B**). The top 10 earliest and latest keywords are summarized in **Table 3**. "Noninvasive monitoring" (blue, occurrences 6, Avg, pub, year: 2008.17) was the earliest co-occurrence keyword, and "extracorporeal membrane oxygenation" (yellow, occurrences 5, Avg, pub, year: 2019.60) was the latest keyword (**Supplementary Figure 2**). "Ventilation distribution" (occurrences, 35; Avg, pub, year, 2012.03) had the highest co-occurrence frequency in the earliest keywords, and "PEEP" (occurrences, 13; Avg, pub, year, 2017.00) had the highest co-occurrence frequency in the latest keywords.

Keyword Burst Value Analysis

The keyword burst value analysis identified the hot spot keywords that have attracted the attention of peer investigators within a certain period. The top 25 keywords with the strongest burst value are summarized in **Figure 4**. During the entire period from 2003 to 2021, "electrical impedance tomography" (strength 7.88) had the highest burst strength, followed by "derecruitment" (strength 7.06), "monitoring" (strength 6.49), "mortality" (strength 6.1), and "functional EIT" (strength 5.55). Separately in 2001–2017, keywords about ventilation distribution monitoring device including "EIT," "derecruitment," "monitoring," "bedside" (strength 5.27), and "spatial distribution" (strength 5.18) were strongly concerned, and in 2017–2021, keywords about ventilation injury and strategy, such as "protective ventilation" (strength 5.37), "titration" (strength 5.46), "transpulmonary pressure" (strength 5.18), "driving pressure" (strength 3.67), and induced "lung injury" (strength 4) were strongly cited. The keywords that still maintained a high brush value until now are "driving pressure," "respiratory failure" (strength 3.66), and "titration" (strength 5.46), which may be the recent hot spot topics of investigators.

Bibliometric Analysis of Co-cited Reference Analysis

Co-cited Reference Clustering and Time Evolution Analysis

The co-cited reference clustering report exhibited a completely mean silhouette of 0.87 and a whole modularity Q score of 0.76, indicating that the clustering effect is efficient and convincing and the features and definition of every subdomain were distinct. Within the analysis, publications on EIT lung monitoring research were divided into 20 clusters (**Figure 5**), a vertically descending order showed cluster size, and cluster labels were obtained using the log-likelihood ratio (LLR) and mutual information (MI). The largest cluster is "obstructive lung diseases" (#0, size: 97) and the smallest cluster is "lung perfusion" (#20, size: 5). The next five largest clusters are "lung collapse" (#1, size: 94), "regional lung volume" (#2, size: 81), "acute hypoxemic respiratory failure" (#3, size: 72), "spontaneous effort" (#4, size: 66), and "increasing PEEP" (#5, size: 52). The large six clusters are summarized in **Supplementary Table 1**. Co-cited reference time evolution analysis shows that in the largest six clusters, "lung collapse" (1#), "regional lung volume change" (2#), and "increasing PEEP" (3#) were highly cited before 2010. "Obstructive lung disease" (0#), "regional lung volume change" (3#), and "acute hypoxemic respiratory failure" (4#) are highly cited since 2010 and kept as citation hotspot until now. Interestingly, despite the Q score, our result showed the two clusters (#6, #11) labeled as "Preterm lambs" using LLR.

Co-cited Reference Burst Value and SVA

The top 25 co-cited references with the strongest burst value on EIT applications in clinical lung monitoring from 2001 to 2020 are summarized in **Figure 6**. Top co-cited references are mostly from the specialty journals of respiratory and critical care, eight animal studies, seven reviews, and 10 clinical studies were included, and references began to burst since 2003, the

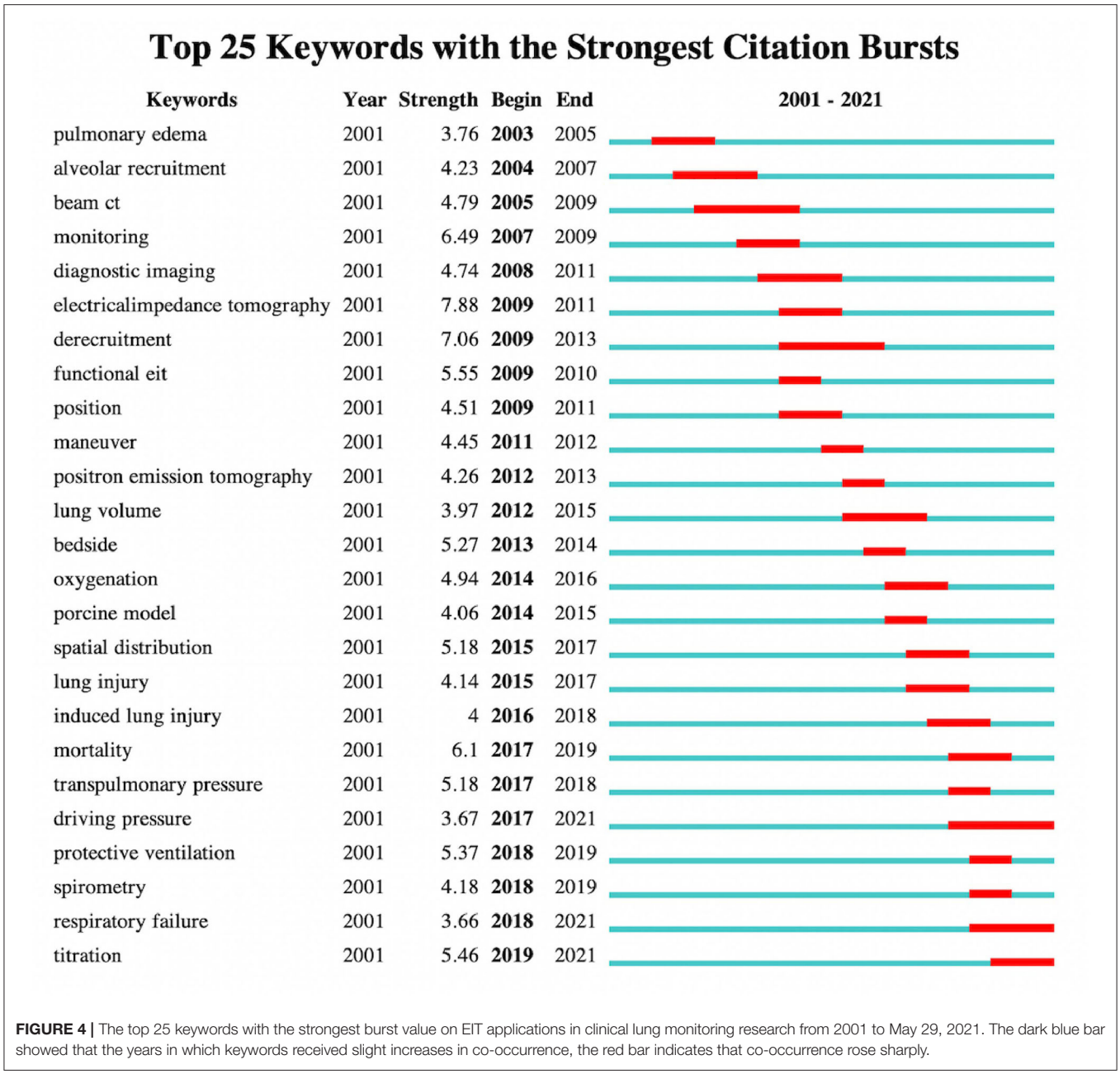


FIGURE 4 | The top 25 keywords with the strongest burst value on EIT applications in clinical lung monitoring research from 2001 to May 29, 2021. The dark blue bar showed that the years in which keywords received slight increases in co-occurrence, the red bar indicates that co-occurrence rose sharply.

publication "Frerichs I, 2017, THORAX" (12) (strength 47.45) had the highest burst strength, next four references with the burst strength above 15 were "Victorino JA, 2004, AM J RESP CRIT CARE" (11) (strength 25.04), "Meier T, 2008, INTENS CARE MED" (16) (strength 20.33), "Frerichs I, 2002, J APPL PHYSIOL" (13) (strength 16.29), and "Costa ELV, 2009, INTENS CARE MED" (21) (strength 15.82). The average duration of the burst value lasts for 3–4 years, "Amato MBP, 2015. NEW ENGL J MED" (22) (strength 11.25, 2016–2021) is the only reference with a burst duration of more than 5 years and is still highly cited today with the other three literature "Frerichs I, 2017, THORAX" (12) (strength 47.45, 2017–2021), "Bellani G, 2016, JAMA-J AM MED ASSOC" (strength 10.2, 2017–2021) (23), and "Spadaro S, 2018, CRIT CARE" (24) (strength 9.31, 2019–2021). Co-cited reference burst value analysis showed the publications that are highly frequently cited in a specific period imply rising research interests of the topic of these publications.

To identify the capacity of cited references to make extraordinary or unexpected connections across distinct clusters and detect the potential landmark studies in the EIT research field, the SVA was conducted. The top five structurally variational references are listed in **Table 4**. The clustering and clustering time evolution situations of these articles are marked in **Figure 7**. The publication with the highest modularity change rate in our

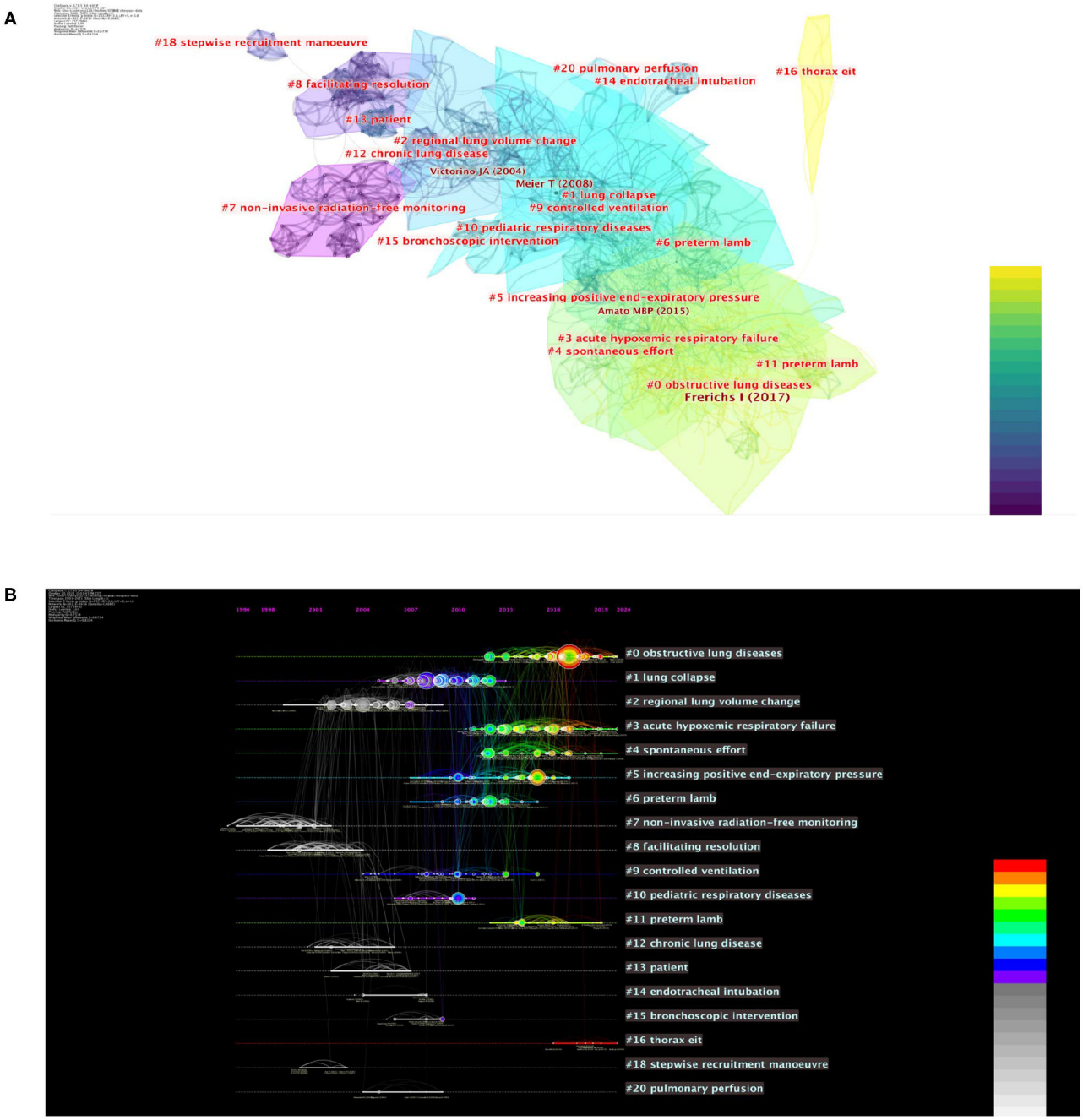


FIGURE 5 | The co-cited reference clustering and time evolution analysis on EIT applications in clinical lung monitoring research from 2001 to May 29, 2021. **(A)** Co-cited reference network clustering analysis. The color indicates the publication year. The descending order vertically indicated cluster size. The line thickness indicated the strength of linkage. **(B)** The co-cited reference time evolution analysis. The color indicates the publication year. The colored curves represent co-citation links added in the year of the corresponding color. The large-sized nodes indicated they are either highly cited or have citation bursts or both. The descending order vertically indicated cluster size. The publication year is indicated above the map.

data set is "Frerichs I, 2017, THORAX" (12) (the modularity change rate: 98.49; cited frequency: 147), spanning three clusters: "obstructive lung diseases" (#1), "acute hypoxemic respiratory failure" (#3), and "spontaneous effort" (#4). "Leonhardt S, 2012, INTENS CARE MED" (26) (the modularity change rate: 95.06; cited frequency: 96); "Adler A, 2012, PHYSIOL MEAS" (25) (the modularity change rate: 93.82; cited frequency: 108) spanning the same clusters "pediatric respiratory diseases" (#10) and #1 "lung collapse" (#1) took the second and third place. Three of the top five structural variational references were review articles, among them "Leonhardt S, 2012, INTENS CARE MED" (26), a review summary of the state-of-the-art in EIT for ventilation

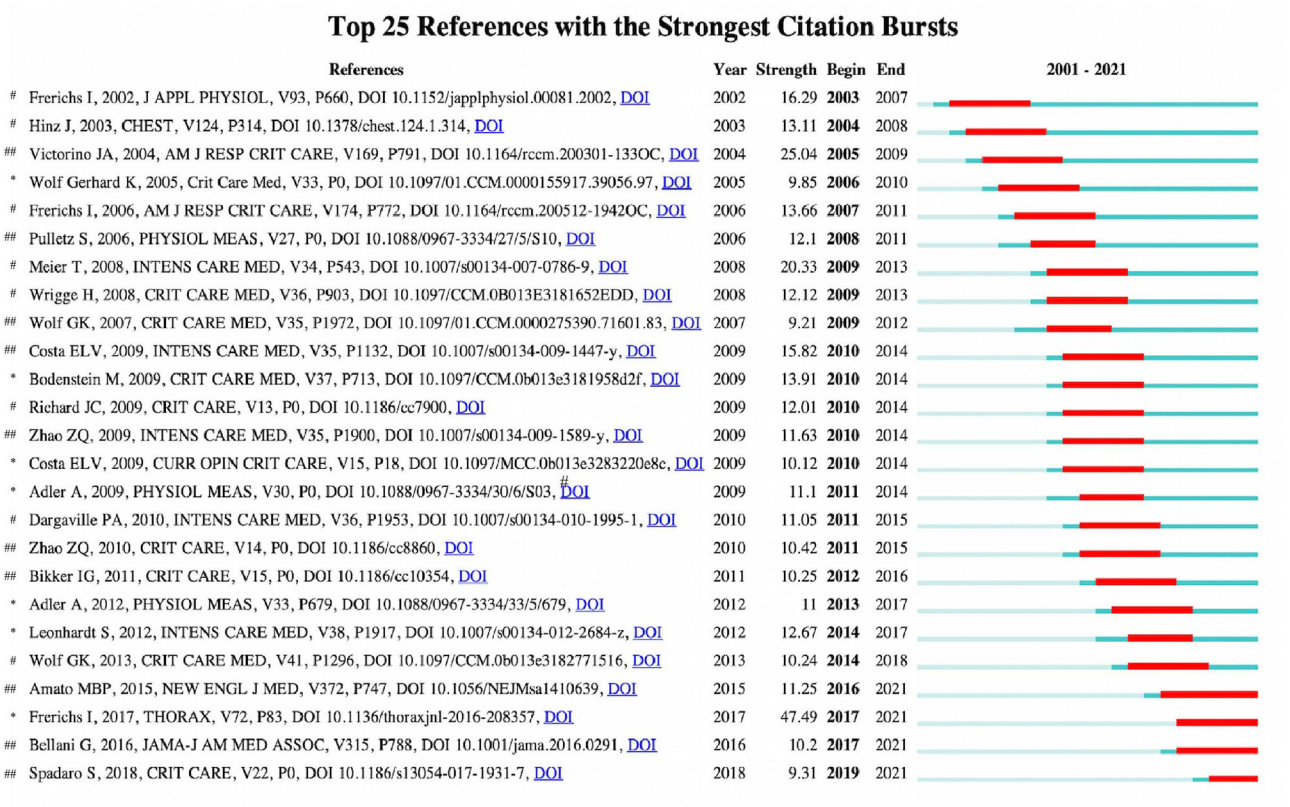


FIGURE 6 | The top 25 cited references with the strongest burst value on EIT applications in clinical lung monitoring research from 2001 to May 29, 2021. The dark blue bar indicated the years in which keywords slightly increased as a co-occurrence. In contrast, the red bar shows that co-occurrence rises sharply. *Review, #Animal research, and ##Clinical research.

TABLE 4 | The top five co-cited references with the strongest structural variation value.

Title	Publication type	Author	Publication year	Journal	Modularity change rate	Citations	Cluster linkage	Centrality divergence
Chest electrical impedance tomography examination data analysis terminology clinical use and recommendations: consensus statement of the TRanslational EIT developmeNt stuDy group*	Review	Frerichs I	2017	Thorax	98.49	147	0.73	0.16
Electrical impedance tomography: the holy grail of ventilation and perfusion monitoring?*	Review	Leonhardt S	2012	Intens Care Med	95.06	96	0.03	0.82
Whither lung EIT: Where are we where do we want to go and what do we need to get there?*	Review	Adler A	2012	Physiol Meas	93.82	108	3.75	0.59
Detection of local lung air content by electrical impedance tomography compared with electron beam CT#	Original Article	Frerichs I	2002	J Appl Physiol	90.68	138	67.04	0.35
Imbalances in regional lung ventilation—A validation study on electrical impedance tomography##	Original Article	Victorino JA	2004	Am J Resp Crit Care	88.57	196	24.94	0.37

*Modularity change rate: the structural changes of the underlying network are induced due to connections with a contribution from new publications. Cluster linkage: an effect of between-cluster links before and after a new paper becomes available. Centrality divergence: the structural variations arising from a new article based on the divergence of the distributions between the central measures of all nodes in the network before and after the information from the new article are considered. *Review, #Animal research, ##Clinical research.*

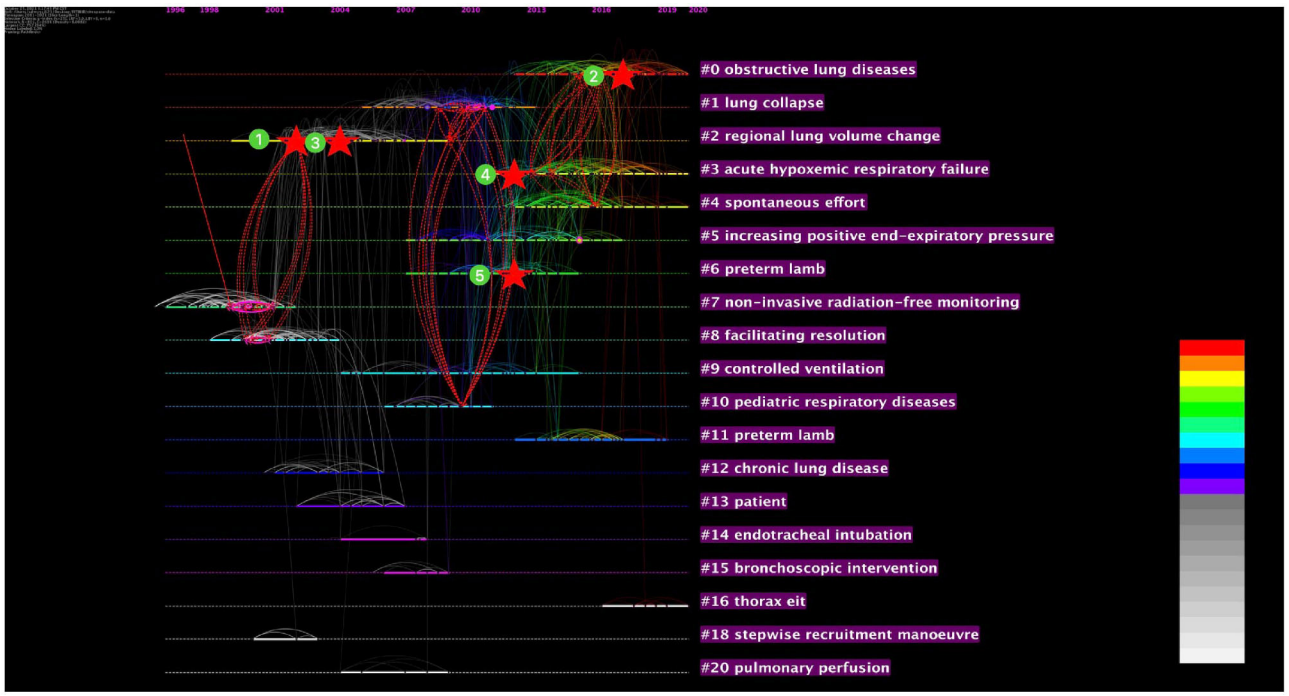


FIGURE 7 | The co-cited reference structural variation analysis (SVA) on EIT lung monitoring research from 2001 to May 29, 2021 (Modularity $Q = 0.76$; Weighted Mean Silhouette $S = 0.87$). The stars indicate citing articles, and the dashed lines indicate novel co-citation links. [①: Victorino et al. (11); ②: Frerichs et al. (12); ③: Frerichs et al. (13); ④: Adler et al. (25); ⑤: Leonhardt et al. (26)].

and perfusion imaging, had the lowest number of citations, but took the second place of the modularity change rate. All the top five structural variation references have a burst value of at least 11. Boundary-spanning ideas may contribute to interdisciplinary scientific and technological progress and raise significant concerns among investigators.

Category Co-occurrence Analysis

The category co-occurrence map showed 62 disciplines cross and penetrate each other with 172 link lines (**Figure 8**). The largest five categories are "GENERAL & INTERNAL MEDICINE" (frequency: 213; Centrality: 0.24), "CRITICAL CARE MEDICINE" (frequency: 192; Centrality: 0.01), "PHYSIOLOGY" (frequency: 136; Centrality: 0.48), "ENGINEERING" (frequency: 136; Centrality: 0.84), and "ENGINEERING, BIOMEDICAL" (frequency: 125; Centrality: 0.24). Among the major categories, "ENGINEERING" with a high centrality of 0.84 showed currently active interdisciplinary crossing, "CRITICAL CARE MEDICINE" with a centrality score of 0.01 represents weak interdisciplinary crossing.

DISCUSSION

This study performs a bibliometric analysis of publications on EIT applications in clinical lung monitoring from 2001 to 2021. A total of 636 publications were analyzed (**Figure 1**), the result reflects the development process, the current situation, emerging trends, and hot spots of the research field over time. The number of publications showed an approximately increasing trend from 2001 to 2020, suggesting that EIT has become a concerned research field (**Figure 2**).

The publication composition indicated that Germany as the pioneer country published the largest number of articles on EIT with lung monitoring every year from 2001 to 2021 (**Figure 2**). It was noticed that there was a remarkable increase in Chinese publications after 2019, this may be because of the availability of EIT equipment for clinical research after obtaining the license from the China Food and Drug Administration in 2014 (1) and increasing demand for bedside non-invasive lung monitoring technology after the outbreak of COVID-19 (27–29).

Citation counts of institutions, journals, authors, and reference articles generally represent scientific acknowledgment for peer investigators (30). Highly cited institutions, authors, and countries cooperated closely and published a considerable number of publications, as shown in **Table 1** and **Supplementary Figure 1**. International scientific cooperation can enhance research quality and promote medical progress (31), follow the trend of international diversification, join online or offline international specialized conferences, and seek interdisciplinary potential and effective ways to promote transnational cooperation. Journal citation analysis can provide reliable references for researchers to search literature or submit manuscripts (32). The journal Physiological Measurement has a tradition of collaborating with the EIT annual conference and publishes a special issue each year after the conference. It produced the highest number of publications and had the

FIGURE 8 | The category co-occurrence analysis on EIT applications in clinical lung monitoring research from 2001 to May 29, 2021. The circle indicates the category; the circle size indicates publication number, the color shows the first publication year. The thickness of the lines indicated the strength of the linkage. The purple ring indicates centrality above 0.1 of categories. Centrality: metric score indicates the catenation value of the category in the entire network structure, ranges from 0–1, above 0.1 indicate turning points.

highest citation. Critical care-specialized journals contribute most publications and cited frequency in the research field as the most promising application direction of lung monitoring with EIT occurs in the ICU. The validation of animal or clinical studies between EIT and gold standards, such as CTs under physiological conditions or disease models (11, 13, 20), studies defining new parameter indicators from EIT (15, 16), and systematic reviews of EIT imaging technology and clinical applications with increasing clinical evidence (12) were highly cited (**Table 1**). All these results indicate that EIT, as a convenient lung function measuring device, plays an essential role in the field of daily and scientific research use in critical care and intensive care medicine.

Refining the research topic with keywords, in the last two decades, the primary research direction of EIT with clinical lung monitoring has focused on diagnostic imaging for ARDS, which is the most life-threatening and difficult respiratory syndrome worldwide and is possible to remain as research hot spots over the next few years. The biggest advantage of EIT technology is the potential to monitor the regional pulmonary function of intensive care patients at the bedside, and early studies focused on clinical use scenarios related to EIT, such as "noninvasive monitoring," "intensive care unit," and monitoring indicators that can be completed, such as "lung impedance" and "ventilation distribution." With the clinical acceptance and promotion of the technology, clinical research of EIT more specifically focuses on optimal respiratory treatment strategies for ARDS, such as monitoring "regional mechanical ventilation distribution" and "PEEP titration" (**Figure 4**, **Tables 2**, **3**), the controversies of protective ventilation in ARDS provided researchers the opportunities to study ARDS using EIT. With more recent experience in clinical use, the topics extend beyond ARDS back to EIT technology as exploring the ventilation statues of new respiratory treatment methods for acute respiratory failure (ARF), e.g., "high-flow nasal cannula" and "extracorporeal membrane oxygenation" as well as a specific population, e.g., "obesity" or subjects under "general anesthesia" become an emerging trend of EIT-related clinical studies (**Table 3**, **Supplementary Figure 2**).

Our keyword analysis also showed the understanding of EIT clinic applications on one topic changing over time, for example, "ARDS," early burst keywords focus on the clinical phenomenon, such as "pulmonary edema." With advances in clinical practice and scientific research, investigators found that the nature

behind these phenomena was "lung injury" and subsequent research gradually focused on clinical intervention strategies, such as "lung protective ventilation." The same holds true for the understanding of "lung recruitment": early understanding may lie in some diseases causing pathophysiological changes as "alveolar collapse" and with the clinical introduction of EIT, it was recognized that such physiological changes can be visualized through diagnostic images as "lung recruitment" and "derecruitment" after which we can apply this topic to clinical treatments, such as "PEEP titration" (**Figure 4**). The depth and breadth of the clinical application of EIT have also changed over time. As mentioned earlier, the current clinical research of EIT is not limited to "ARDS" but an evolution to more generalized "ARF," the Research Topic expands beyond the ventilation, such as "perfusion" and the clinical application scenario is not limited to the intensive care unit but anesthesia. These changes interact with the discipline's cognitive progress of disease and the clinical scientific research application of visual EIT monitoring.

References are essential for the selection, execution, and summary of scientific research. The high-quality literature of a specific research object can give a reference for this research direction and form a system through positive research accumulation. References of publications on EIT applications in clinical lung monitoring cluster toward the evaluation of ARDS lung ventilation like lung collapse and increasing PEEP in the early years (**Figure 5**). Presently, similar to keyword analysis, respiratory failure, regional ventilation, and extensive pulmonary function monitoring may become research hot spots in EIT clinical use. From our results, for topics with a high clustering intensity at an early age but not in recent years, e.g., "lung perfusion," technical bottleneck, and clinical adaptation might be the reasons for changes.

Focus on the content of cited reference: Among studies with high burst values, in early days, animal studies that validated EIT and imaging gold standards (13, 18), and animal model studies (16, 20) had high burst values, as medium animals, such as pigs and sheep, are preferred for simulating respiratory physiology and establishing disease model for validation studies. This is vital for introducing EIT as a new technology for clinical ventilation monitoring. Since EIT was used in complex clinical scenarios, volunteer and patient studies (11, 21) are frequently cited. In addition, clinical studies introduced new parameters calculated by EIT (15, 33) and related articles in ARDS treatment (34) are also highly cited (**Figure 6**). Another highly cited type is reviews. The topic of EIT-related review started on the principles and EIT indications, and then focused on the summary of clinical applications, the exploration of new fields, "Frerichs I, 2017, THORAX" (12), a consensus statement of the translational EIT development study group provides examination, data analysis, terminology, and clinical use recommendations of EIT, had the highest burst strength three times higher than others. The literature "Amato MBP, 2015, NEW ENGL J MED" (22), an observational trial regarding driving pressure in ARDS without mentioning EIT, is the only reference with a burst duration of more than 5 years, and is still highly cited today (**Figure 6**). It was supposed that this article brings the investigators to the physiological changes of local ventilation during ARDS, which is exactly what EIT monitoring technology can approve. Our analysis suggests that a high-quality review or guideline from pioneers in the professional field may provide reliable evidence for the beginning of new studies on EIT related to clinical lung monitoring. Additionally, the progress and concept of clinical diseases, implying rising research interest trends, were also essential for promoting EIT's clinical use. As a visualized ventilation monitoring tool, a gradual understanding of EIT and its application need multiple levels of validation. with a deeper understanding of disease and EIT technology, the breadth of future researches will be expanded.

The structural variation analysis was conducted to find the co-cited reference spanning cluster boundaries. High structural variation references are high burst value reviews (**Figure 7** and **Table 4**). "Frerichs I, 2017, THORAX (12)," a clinical application consensus gives a detailed clinical application protocol of EIT, spanning three clusters. "Leonhardt S, 2012, INTENS CARE MED (26)" reviewed EIT applications in ventilation and perfusion imaging and got a high modularity change rate relative to cited frequency. That is probably because it elaborated special respiratory monitoring using EIT, which gives an impetus to cross-category research. It is not difficult to infer that macro-overview or review depth in branch direction have a high reference value for research on EIT monitoring.

Some studies on EIT applications in clinical lung monitoring have the characteristics of interdisciplinary crossing. "PHYSIOLOGY," "ENGINEERING," and "RESPIRATORY SYSTEM," the three categories, represent the technical foundation and operational value of EIT, which confirmed the active interdisciplinary crossing. "CRITICAL CARE MEDICINE" is a hot subject, but building a collaboration with other disciplines needs to be addressed in further research (**Figure 8**).

STRENGTHS AND LIMITATIONS

This study is the first bibliometric analysis evaluating publications on the application of EIT in clinical lung monitoring extracted from the WoS core database. This study provides a quick and objective reference for interested researchers by visualizing the current status, hot spots, and emerging trends of EIT from 2001 to 2021. However, some limitations are inevitable. First, the WoSCC database is updated continuously and dynamically. There might be some new data missing, even all the database searches were conducted in 1 day. Second, to obtain more subject-oriented research, only English original articles and reviews about clinical applications were included, the sample size of articles finally included in the analysis is limited and a discrepancy may exist between our results and the real publication characteristics. Finally, the multiple expressions of author, institution, and keywords result in the dispersion of counts and clusters. Although these problems were addressed with the merge and normalization function of the software, they cannot be avoided completely.

CONCLUSION

Electrical impedance tomography applications in clinical lung monitoring are concerned research fields from 2001 to 2021. Germany was a pioneer country in this research field, while the Univ Med Ctr Schleswig and Frerichs I achieved significant research results and contributed to the development of EIT research. Professional macro and in-depth review and interdisciplinary literature could give a reliable reference for EIT research, while cooperation would promote the development of the research field. Ventilation distribution in ARDS and respiratory therapy strategies were research focus in the past two decades and will continue as research hot spots. More diversified lung function monitoring techniques, such as lung perfusion and interdisciplinary crossing with EIT, are potential emerging research trends on EIT applications in clinical lung monitoring.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: WoS database, http://isiknowledge.com/.

AUTHOR CONTRIBUTIONS

ZL, SQ, and ZZ have designed the study and drafted the manuscript. YY and SM have performed collected the data. ZL, SQ, and CC have analyzed the data. WL and YG have revised the manuscript. All authors have approved the final version.

FUNDING

This work was supported by the Shanghai Jiao Tong University (No. YG2019ZDB04), Natural Science Research Project of Minhang District (No. 2019MHZ017), and Shanghai Renji Hospital Clinical Research and Cultivation Fund (No. PY2018-IIA-01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.813640/full#supplementary-material

REFERENCES

1. Zhao Z, Fu F, Frerichs I. Thoracic electrical impedance tomography in Chinese hospitals: a review of clinical research and daily applications. *Physiol Meas.* (2020) 41:04TR01. doi: 10.1088/1361-6579/ab81df
2. Lowhagen K, Lindgren S, Odenstedt H, Stenqvist O, Lundin S. A new non-radiological method to assess potential lung recruitability: a pilot study in ALI patients. *Acta Anaesthesiol Scand.* (2011) 55:165–74. doi: 10.1111/j.1399-6576.2010.02331.x
3. Putensen C, Wrigge H, Zinserling J. Electrical impedance tomography guided ventilation therapy. *Curr Opin Crit Care.* (2007) 13:344–50. doi: 10.1097/MCC.0b013e328136c1e2
4. van Genderingen HR, van Vught AJ, Jansen JR. Estimation of regional lung volume changes by electrical impedance pressures tomography during a pressure-volume maneuver. *Intensive Care Med.* (2003) 29:233–40. doi: 10.1007/s00134-002-1586-x
5. Yun L, He HW, Moller K, Frerichs I, Liu D, Zhao Z. Assessment of lung recruitment by electrical impedance tomography and oxygenation in ARDS patients. *Medicine.* (2016) 95:e3820. doi: 10.1097/MD.0000000000003820
6. Zhao Z, Chang MY, Chang MY, Gow CH, Zhang JH, Hsu YL, et al. Positive end-expiratory pressure titration with electrical impedance tomography and pressure-volume curve in severe acute respiratory distress syndrome. *Ann Intensive Care.* (2019) 9:7. doi: 10.1186/s13613-019-0484-0
7. Wiysonge CS, Uthman OA, Ndumbe PM, Hussey GD. A bibliometric analysis of childhood immunization research productivity in Africa since the onset of the Expanded Program on Immunization in 1974. *BMC Med.* (2013) 11:66. doi: 10.1186/1741-7015-11-66
8. Turatto F, Mazzalai E, Pagano F, Migliara G, Villari P, De Vito C. A systematic review and bibliometric analysis of the scientific literature on the early phase of COVID-19 in Italy. *Front Public Health.* (2021) 9:666669. doi: 10.3389/fpubh.2021.666669
9. Pirri S, Lorenzoni V, Turchetti G. Scoping review and bibliometric analysis of Big Data applications for Medication adherence: an explorative methodological study to enhance consistency in literature. *BMC Health Serv Res.* (2020) 20:688. doi: 10.1186/s12913-020-05544-4
10. Liang YD, Li Y, Zhao J, Wang XY, Zhu HZ, Chen XH. Study of acupuncture for low back pain in recent 20 years: a bibliometric analysis via CiteSpace. *J Pain Res.* (2017) 10:951–64. doi: 10.2147/JPR.S132808
11. Victorino JA, Borges JB, Okamoto VN, Matos GF, Tucci MR, Caramez MP, et al. Imbalances in regional lung ventilation: a validation study on electrical impedance tomography. *Am J Respir Crit Care Med.* (2004) 169:791–800. doi: 10.1164/rccm.200301-133OC
12. Frerichs I, Amato MB, van Kaam AH, Tingay DG, Zhao Z, Grychtol B, et al. Chest electrical impedance tomography examination, data analysis, terminology, clinical use and recommendations: consensus statement of the TRanslational EIT developmeNt stuDy group. *Thorax.* (2017) 72:83–93. doi: 10.1136/thoraxjnl-2016-208357
13. Frerichs I, Hinz J, Herrmann P, Weisser G, Hahn G, Dudykevych T, et al. Detection of local lung air content by electrical impedance tomography compared with electron beam CT. *J Appl Physiol.* (2002) 93:660–6. doi: 10.1152/japplphysiol.00081.2002
14. Adler A, Arnold JH, Bayford R, Borsic A, Brown B, Dixon P, et al. GREIT: a unified approach to 2D linear EIT reconstruction of lung images. *Physiol Meas.* (2009) 30:S35–55. doi: 10.1088/0967-3334/30/6/S03
15. Zhao Z, Moller K, Steinmann D, Frerichs I, Guttmann J. Evaluation of an electrical impedance tomography-based Global Inhomogeneity Index for pulmonary ventilation distribution. *Intensive Care Med.* (2009) 35:1900–6. doi: 10.1007/s00134-009-1589-y
16. Meier T, Luepschen H, Karsten J, Leibecke T, Grossherr M, Gehring H, et al. Assessment of regional lung recruitment and derecruitment during a PEEP trial based on electrical impedance tomography. *Intensive Care Med.* (2008) 34:543–50. doi: 10.1007/s00134-007-0786-9
17. Frerichs I. Electrical impedance tomography (EIT) in applications related to lung and ventilation: review of experimental and clinical activities. *Physiol Meas.* (2000) 21:R1–21. doi: 10.1088/0967-3334/21/2/201
18. Hinz J, Neumann P, Dudykevych T, Andersson LG, Wrigge H, Burchardi H, et al. Regional ventilation by electrical impedance tomography: a comparison with ventilation scintigraphy in pigs. *Chest.* (2003) 124:314–22. doi: 10.1378/chest.124.1.314
19. Brower RG, Matthay MA, Morris A, Schoenfeld D, Thompson BT, Wheeler A. Ventilation with lower tidal volumes as compared with traditional tidal volumes for acute lung injury and the acute respiratory distress syndrome. *N Engl J Med.* (2000) 342:1301–08. doi: 10.1056/NEJM200005043421801
20. Frerichs I, Dargaville PA, van Genderingen H, Morel DR, Rimensberger PC. Lung volume recruitment after surfactant administration modifies spatial distribution of ventilation. *Am J Respir Crit Care Med.* (2006) 174:772–9. doi: 10.1164/rccm.200512-1942OC
21. Costa ELV, Lima RG, Amato MBP. Electrical impedance tomography. *Curr Opin Crit Care.* (2009) 15:18–24. doi: 10.1097/MCC.0b013e3283220e8c

22. Amato MBP, Meade MO, Slutsky AS, Brochard L, Costa ELV, Schoenfeld DA, et al. Driving pressure and survival in the acute respiratory distress syndrome. *N Engl J Med*. (2015) 372:747–55. doi: 10.1056/NEJMsa1410639
23. Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA*. (2016) 315:788–800. doi: 10.1001/jama.2016.0291
24. Spadaro S, Mauri T, Bohm SH, Scaramuzzo G, Turrini C, Waldmann AD, et al. Variation of poorly ventilated lung units (silent spaces) measured by electrical impedance tomography to dynamically assess recruitment. *Crit Care*. (2018) 22:26. doi: 10.1186/s13054-017-1931-7
25. Adler A, Amato MB, Arnold JH, Bayford R, Bodenstein M, Bohm SH, et al. Whither lung EIT: where are we, where do we want to go and what do we need to get there? *Physiol Meas*. (2012) 33:679–94. doi: 10.1088/0967-3334/33/5/679
26. Leonhardt S, Lachmann B. Electrical impedance tomography: the holy grail of ventilation and perfusion monitoring? *Intensive Care Med*. (2012) 38:1917–29. doi: 10.1007/s00134-012-2684-z
27. Sella N, Zarantonello F, Andreatta G, Gagliardi V, Boscolo A, Navalesi P. Positive end-expiratory pressure titration in COVID-19 acute respiratory failure: electrical impedance tomography vs. PEEP/FiO2 tables. *Crit Care*. (2020) 24:540. doi: 10.1186/s13054-020-03242-5
28. Perier F, Tuffet S, Maraffi T, Alcala G, Victor M, Haudebourg AF, et al. Electrical impedance tomography to titrate positive end-expiratory pressure in COVID-19 acute respiratory distress syndrome. *Crit Care*. (2020) 24:678. doi: 10.1186/s13054-020-03414-3
29. Spinelli E, Mauri T, Fogagnolo A, Scaramuzzo G, Rundo A, Grieco DL, et al. Electrical impedance tomography in perioperative medicine: careful respiratory monitoring for tailored interventions. *BMC Anesthesiol*. (2019) 19:140. doi: 10.1186/s12871-019-0814-7
30. Zhou S, Tao Z, Zhu Y, Tao L. Mapping theme trends and recognizing hot spots in postmenopausal osteoporosis research: a bibliometric analysis. *PeerJ*. (2019) 7:e8145. doi: 10.7717/peerj.8145
31. Cooper ID. Bibliometrics basics. *J Med Libr Assoc*. (2015) 103:217–8. doi: 10.3163/1536-5050.103.4.013
32. Oelrich B, Peters R, Jung K. A bibliometric evaluation of publications in urological journals among European Union countries between 2000–2005. *Eur Urol*. (2007) 52:1238–48. doi: 10.1016/j.eururo.2007.06.050
33. Pulletz S, van Genderingen HR, Schmitz G, Zick G, Schadler D, Scholz J, et al. Comparison of different methods to define regions of interest for evaluation of regional lung ventilation by EIT. *Physiol Meas*. (2006) 27:S115–27. doi: 10.1088/0967-3334/27/5/S10
34. Zhao Z, Steinmann D, Frerichs I, Guttmann J, Moller K. PEEP titration guided by ventilation homogeneity: a feasibility study using electrical impedance tomography. *Crit Care*. (2010) 14:R8. doi: 10.1186/cc8860

Conflict of Interest: ZZ receives a consulting fee from Dräger Medical.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Medicine

PERSPECTIVE
published: 08 February 2022
doi: 10.3389/fmed.2021.793815

Transportability and Implementation Challenges of Early Warning Scores for Septic Shock in the ICU: A Perspective on the TREWScore

*Michael S. A. Niemantsverdriet[1,2], Meri R. J. Varkila[3], Jacqueline L. P. Vromen-Wijsman[3], Imo E. Hoefer[1], Domenico Bellomo[2], Martin H. van Vliet[2], Wouter W. van Solinge[1], Olaf L. Cremer[3] and Saskia Haitjema[1]**

[1] Central Diagnostic Laboratory, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands, [2] SkylineDx, Rotterdam, Netherlands, [3] Department of Intensive Care Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands

The increased use of electronic health records (EHRs) has improved the availability of routine care data for medical research. Combined with machine learning techniques this has spurred the development of early warning scores (EWSs) in hospitals worldwide. EWSs are commonly used in the hospital where they have been developed, yet few have been transported to external settings and/or internationally. In this perspective, we describe our experiences in implementing the TREWScore, a septic shock EWS, and the transportability challenges regarding domain, predictors, and clinical outcome we faced. We used data of 53,330 ICU stays from Medical Information Mart for Intensive Care-III (MIMIC-III) and 18,013 ICU stays from the University Medical Center (UMC) Utrecht, including 17,023 (31.9%) and 2,557 (14.2%) cases of sepsis, respectively. The MIMIC-III and UMC populations differed significantly regarding the length of stay (6.9 vs. 9.0 days) and hospital mortality (11.6% vs. 13.6%). We mapped all 54 TREWScore predictors to the UMC database: 31 were readily available, seven required unit conversion, 14 had to be engineered, one predictor required text mining, and one predictor could not be mapped. Lastly, we classified sepsis cases for septic shock using the sepsis-2 criteria. Septic shock populations (UMC 31.3% and MIMIC-III 23.3%) and time to shock events showed significant differences between the two cohorts. In conclusion, we identified challenges to transportability and implementation regarding domain, predictors, and clinical outcome when transporting EWS between hospitals across two continents. These challenges need to be systematically addressed to improve model transportability between centers and unlock the potential clinical utility of EWS.

Keywords: early warning score (EWS), TREWScore, sepsis, septic shock, intensive care

OPEN ACCESS

Edited by:
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

Reviewed by:
Yong Ming Yao,
First Affiliated Hospital of Chinese PLA General Hospital, China
Qilin Yang,
The Second Affiliated Hospital of Guangzhou Medical University, China

****Correspondence:***
Saskia Haitjema
S.Haitjema@umcutrecht.nl

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *12 October 2021*
Accepted: *22 December 2021*
Published: *08 February 2022*

Citation:
Niemantsverdriet MSA, Varkila MRJ, Vromen-Wijsman JLP, Hoefer IE, Bellomo D, van Vliet MH, van Solinge WW, Cremer OL and Haitjema S (2022) Transportability and Implementation Challenges of Early Warning Scores for Septic Shock in the ICU: A Perspective on the TREWScore. Front. Med. 8:793815. doi: 10.3389/fmed.2021.793815

INTRODUCTION

Early recognition and diagnosis of hospitalized patients at risk of clinical deterioration is crucial for adequate treatment. To aid healthcare professionals in their systematic assessment of these patients, a plethora of flowcharts and early warning scores (EWSs) have been developed for various diseases. Such scores are especially relevant for early detection of potentially life-threatening syndromes

where time is of essence (1–3). For sepsis, in particular, several EWS have been developed for the identification of preseptic patients in hospital wards, emergency departments (EDs), and intensive care units (ICUs) (4). The continuous registration and collection of clinical and vital parameters in the ICU provide a unique opportunity to continuously calculate EWS for sepsis and adjust treatment thereupon (5). Consequently, many such scores have been developed and published, albeit almost exclusively in single centers. As EWS are currently not transported and implemented to other centers, the universal potential of these models cannot be fully exploited yet.

Transporting scores between healthcare settings faces several challenges. First, the EWS needs to be openly accessible to reproduce the model. Code is often not publicly shared, including the predictors of EWS and, therefore, either needs to be requested or reverse-engineered (5). Second, no hospital is alike as patient populations tend to be rather hospital-specific based on the size of the hospital, location, and case-mix (4). Moreover, EWS input parameters are not always readily available throughout different hospitals as they are either recorded differently or may not be available at all. Finally, EWS have to be carefully tested to assess their clinical validity in the recipient center, a step that is frequently omitted (4). Broad, international, and interhospital application of algorithm-based EWS will have to tackle and overcome these hurdles.

Even though these challenges are widely recognized in literature, we were interested in addressing and evaluating the scope of these challenges related to the domain, predictor, and clinical outcome (6, 7). As a use case, we attempted to transport the Targeted Real-time Early Warning Score (TREWScore) algorithm proposed by Henry et al. in 2015 to our center (8). The TREWScore was trained on electronic health record (EHR) data from the publicly available Medical Information Mart for Intensive Care (MIMIC)-II database to prospectively identify patients with septic shock in the ICU (9). With a high accuracy [0.83 area under the curve (AUC)] at a median diagnostic lead time of 28.2 h before shock onset, the TREWScore was received with great enthusiasm.

We explored transportability and implementation challenges relating to domain, predictors, and clinical outcomes. This study was performed according to the Declaration of Helsinki, the GDPR and the institutional review board approved of the study (registration number 19/543). Only pseudonymized data were used.

Domain Challenge

First, we evaluated the domain by comparing the ICU populations of the MIMIC-III and the University Medical Center (UMC ICU) Utrecht. We included 55,330 and 18,013 consecutive ICU stays for MIMIC-III and UMC ICU, respectively. The MIMIC-III is a publicly available database comprising data from the ICU units of the Beth Israel Deaconess Medical Center collected between 2001 and 2012 (10). UMC Utrecht is a large tertiary referral hospital located in Utrecht, Netherlands. From the UMC, ICU included all patients between 2011 and 2019. From both databases, we included consecutive patients older than 18 years, and data were combined for patients who were readmitted to the ICU within 24 h.

The UMC ICU cohort was younger (64.1 vs. 65.8 years) with a higher proportion of men (56.6 vs. 53.3%) (**Table 1**). ICU length of stay was shorter in the UMC ICU cohort (1.0 vs. 2.2 days), whereas total hospital length of stay was longer in the UMC cohort (9.0 vs. 6.91). Proportions of hospital mortality, blood pressure monitoring, and mechanical ventilation were all higher in the UMC ICU cohort, whereas MIMIC-III had a two-fold higher sepsis prevalence compared to the UMC ICU, 31.8 and 14.2%, respectively. There were thus significant differences in cohort characteristics between MIMIC-III and UMC ICU, with the latter appearing more severely ill.

Predictor Challenge

From a total of 54 predictors, the TREWScore automatically selected 26 by removing uninformative predictors with lasso regularization (11). We mapped all 54 TREWScore predictors used for training to our database. As clinical practice differs among centers, not all predictors were recorded for the UMC ICU cohort and/or were measured in a different unit of measure. Apart from missingness and unit discrepancies, TREWScore also comprises engineered predictors based on a combination of predictors and/or International Classification of Diseases (ICD)-9 codes. As ICD-9 codes are not available in the stored routine care data in the UMC ICU cohort, we extracted these predictors from the Dutch National Intensive Care Evaluation (NICE) minimal dataset, the Dutch ICU Quality Registry that is manually maintained regularly next to the hospital information system. To explore the predictor mapping discrepancy between both cohorts, we followed a staged approach for each predictor: (1) first we checked whether a predictor is part of routine care data, i.e., is it automatically processed and displayed in the EHR system; (2) then we checked if the predictor is available in the EHR system, does it need to be converted to a different unit (3), for predictors not available in the system, we checked whether the predictor components are available, and if so, (4) whether the predictor needed to be engineered or mined from the text.

Results on the TREWScore predictor mapping are shown in **Figure 1** and **Supplementary Table 1**. Of the 54 predictors, 38 predictors were available in the UMC EHR system; 31 were readily available and seven required unit conversion: FiO_2 and hematocrit from percentage to fraction, the hemoglobin from mmol/l to g/dl, admission weight and current weight from kg to pounds, blood urea nitrogen from mmol/l to mg/dl, and serum creatinine from μmol/l to mg/dl. Of the 31 readily available features, eight were based on ICD-9 codes for which we could find a surrogate in the NICE dataset. The remaining 16 predictors required other sources of information in terms of predictor engineering or text mining. The time since the first organ dysfunction (chronic or acute) predictor could not be mapped as the time of organ dysfunction was not clearly defined by Henry et al. (8). The remaining 53 predictors could be mapped by either engineering ($N = 14$) or text mining ($n = 1$). After unit conversion, all predictors were available to engineer the 14 predictors.

TABLE 1 | Characteristics of intensive care unit stays.

	MIMIC-III (*N* = 53,330)	Of which septic shock (*N* = 4,631)	UMC ICU (*N* = 18,013)	Of which septic shock (*N* = 794)
Included years of ICU admission	01-01-2001 / 31-12-2012		01-01-2011 / 30-06-2019	
Distinct patients, count	38,511		17,038	
Hospital admissions, count	49,694		17,195	
Patient characteristics				
Age, years, median [Q1–Q3]	65.8 [52.9–77.9]	67.4 [55.4–79.3]	64.1 [53.0–72.5]	62.0 [52.0–69.0]
Gender, male ICU stays (%)	21,796 (56.6%)	2,469 (53.3%)	11,522 (64.0%)	502 (63.2%)
ICU admissions with at least one sepsis episode during ICU stay, count	17,032 (31.8%)		2,557 (14.2%)	
Characteristics and outcomes of ICU stay				
ICU length of stay, median days [Q1–Q3]	2.2 [1.2–4.2]		1.0 [0.8–3.2]	
Hospital length of stay, median days [Q1–Q3]	6.91 [4.0–11.9]		9.0 [5.8–18.2]	
ICU mortality (%)	4,560 (8.6%)		1,623 (9.0%)	
Hospital mortality (%)	5,739 (11.6%)		2,450 (13.6%)	
Invasive arterial blood pressure monitoring, count (%)	39,149 (73.4%)		17,457 (96.9%)	
Mechanical ventilation, count (%)	25,740 (48.3%)		15,549 (86.3%)	
Length of stay of ICU admissions with at least one septic shock period during ICU stay, count Mean		151.3 [68.9–319.8] 244.2		248.8 [97.1–522.7] 423.4
Time to first shock event, median hours [Q1–Q3] Mean		19.6 [8.7–45.7] 53.2		44.4 [21.6–136.3] 138.7

ICU, Intensive Care Unit.

As real-time algorithms require a continuous feed of data to make predictions, we were interested in the data characteristics in both cohorts in terms of data availability and sampling frequency. As only a subset of the predictors was available in both cohorts, we only included the 22 numerical predictors required to apply the sepsis criteria that were available in both cohorts for this analysis. We assessed the data availability by counting the number of patients with at least one predictor measurement. Around 50% of all MIMIC-III patients had at least one measurement on a majority of the predictors (**Supplementary Figure 1**), whereas over 85% of the UMC ICU patients had at least one measurement available of all predictors.

The sampling frequency was assessed by calculating the average time between predictor measurements for patients with at least two measurements (**Supplementary Table 2**). Continuous vital predictors, i.e., DBP, FiO_2, HR, MBP, RR, SBP, SpO_2, and temperature were registered at a higher continuous rate in the UMC ICU than in the MIMIC-III database, whereas non-vital predictors (e.g., laboratory values) were available in similar sampling times in both cohorts (**Supplementary Figure 2**). Also, we computed the average of each predictor to compare differences in predictor population means between both cohorts. Sampling times were different and the population means between both cohorts. Population summary statistics of the predictors were similar in both cohorts (**Supplementary Figure 3**).

Clinical Outcome Challenge

Following the methodology of Henry et al. (8), we applied the sepsis-2 criteria to identify patients progressing to the clinical outcome of shock in both cohorts (**Supplementary Material Methods**). From the 2,557 sepsis cases in the UMC cohort, 795 (31.3%) developed shock, whereas 3,961 (23.3%) of the 17,023 MIMIC-III sepsis cases developed shock (**Table 1**).

Furthermore, we computed the number of hours in ICU until shock onset, thereby evaluating the potential clinical value of implementing such a model in our center. The majority of MIMIC-III shock patients developed shock in the first 50 h of IC stay (**Supplementary Figure 4**). Median time to first shock event was higher for UMC ICU shock patients compared to MIMIC-III, 44.4 [IQR 21.6–136.3] vs. 19.6 [IQR 8.7–45.7] h, respectively. These findings underline differences in clinical outcomes between MIMIC-III and UMC ICU patients.

DISCUSSION

The transportability of algorithms built on routinely collected EHR could provide hospitals with an early warning for disease. However, here we identified three challenges relating to domain, predictors, and clinical outcome that we encountered when attempting to transport the TREWScore to our center. First, comparisons between the MIMIC-III and UMC ICU cohorts showed differences in medical severity. Second, more than two-thirds (38 out of 54) of the predictors could be readily mapped to the EHR system of our center, one quarter (15 out of 54) of the predictors would have to be compiled with either engineering or text mining and we found differences in data collection and predictor statistics between both centers. Third, the incidence of sepsis and the time to shock were different between both cohorts.

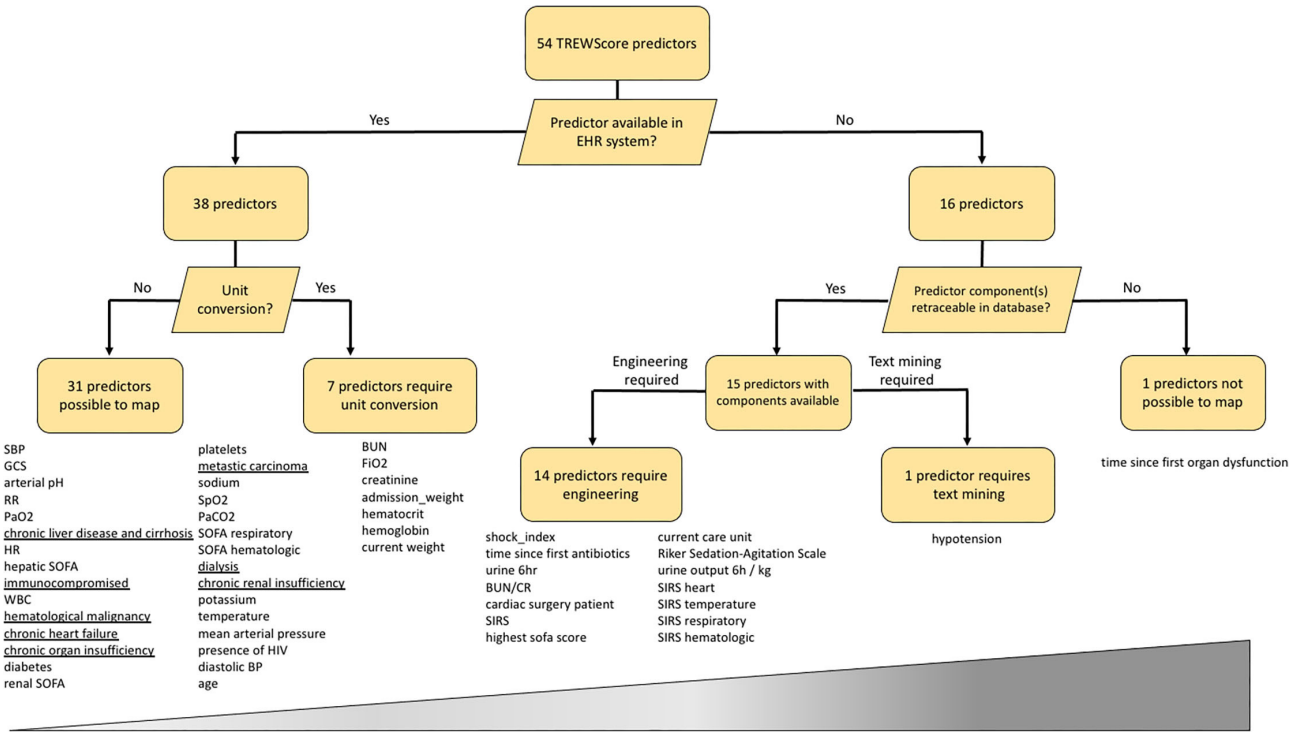


FIGURE 1 | Mapping of the 54 TREWScore predictors to the UMC ICU database. Predictors are listed below each end-node. The increasing width of the bar represents the difficulty scale from easy to hard of mapping each predictor category. Underlined predictors are within 24 h documented in the UMC ICU EHR system and therefore not readily available at ICU admission. EHR, Electronic Health Record; SBP, Systolic Blood Pressure; GCS, Glasgow Coma Scale; RR, Respiratory Rate; HR, Heart Rate; SOFA, Sequential Organ Failure Assessment; WBC, White Blood Cell count; MAP, Mean Arterial Pressure; BP, Blood Pressure; BUN, Blood Urea Nitrogen; BUN/CR, BUN Creatinine Ratio; SIRS, Severe Inflammatory Response Syndrome; HIV, Human Immunodeficiency Virus.

These major challenges in all three domains provide evidence that the TREWScore cannot be easily transported to other centers and question the transportability of EWS in general.

Comparisons between the MIMIC-III and UMC ICU cohorts showed that UMC ICU patients are more intensively monitored in terms of measurement frequency and data availability. This illustrates a difference in clinical practice between a Dutch and a US American center as the ICU in the Netherlands is reserved for critically ill patients in a more acute state, while the American hospitals have a higher proportion of ICU beds in comparison to other countries (12). Nonetheless, sepsis prevalence in the MIMIC-III cohort was significantly higher than in the UMC ICU. Patients in the UMC ICU were evaluated daily clinically for having sepsis as part of the NICE minimal dataset in the Netherlands, whereas MIMIC-III patients were identified as having sepsis based on suspicion of infection defined by ICD-9 codes. As ICD-9 codes may suffer from administrative errors and retrospective allocation, this may have resulted in an overestimation of the sepsis incidence in MIMIC-III.

Because of the identified transportability challenges in this case study, we decided neither to try and validate TREWScore in our center nor further prepare it for implementation. First, TREWScore would require us to alter our clinical workflow to, for example, engineer and collect predictors that are not part of our routine clinical practice, which we would also need to obtain over a longer period to facilitate calibration of the model on existing data from our center. For example, comorbidities are available in the NICE dataset and can be used as surrogates for the ICD-9 codes. However, the physician is requested to document these comorbidities within 24 h after ICU admission, making these data not available at admission. In comparison to the drawbacks of ICD-9 documentation as mentioned above, this system would provide more accurate data for the score. Moreover, we found differences in data collection in terms of frequency of measurements. Lastly, it remains unclear to us how a range of laboratory variables was exactly measured. For example, whether bilirubin was measured conjugated or unconjugated, on what analyzer from what manufacturer using what assay type. Because of all limitations described in our article, we question whether the limited value for our center would outweigh our efforts to implement such a system.

Interestingly, the septic shock time to event was higher for the UMC ICU cohort in comparison with the MIMIC-III cohort. This effect can partly be explained by the longer length of stay of UMC ICU patients as compared to the MIMIC-III cohort. Moreover, the different subgroups of patients in both ICU cohorts could also explain this effect, as explained above. Furthermore, we have not stratified community vs. nosocomial onset and were not able to compare care severity between cohorts as these data were not available. These differences

should be carefully explored before implementing EWS scores in clinical practice.

Data-driven ICU models have the potential to improve patient care in other centers, beyond yielding interesting research papers for literature. Here, we showed compelling evidence that there are many differences between centers in terms of domain, predictors, and clinical outcome. Our perspective resonates with recently reported challenges during external validation efforts of the Epic Sepsis Model for sepsis (13). Moreover, the lack of methodological details to reproduce research and differences in clinical practice currently further complicates transportability.

Data scientists in healthcare that make a commitment to improving patient care have the obligation to sufficiently address transportability issues. First, this means that code and data are made available in an understandable way and that description of methodology facilitates reproducibility, e.g., using current reporting guidelines (14, 15). Second, to improve the predictor mapping, researchers should adhere to data standards, such as FHIR and SNOMED CT, which should be facilitated by IT infrastructures accordingly (16, 17). Third, to further promote uptake around the globe, researchers should use predictors that are ubiquitously available in the medical domain when making models to reduce center and/or region-specific bias. These efforts should help to ameliorate the transportability of models to other centers for external validation and assessment of clinical relevance (18). Here, we show that all three transportability challenges, regarding domain, predictors, and clinical outcome, should be addressed before an EWS can be transported and used in another center. Only then, the true potential of the universal implementation of machine learning models in the intensive care can be assessed and potentially achieved for the benefit of our patients.

DATA AVAILABILITY STATEMENT

Data from the MIMIC-III version 1.4 database is publicly available on (https://mimic.physionet.org). Data from the UMC ICU is available upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the Utrecht Medical University Center. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

WS, OC, and SH conceived the idea of the study and designed the study. MRJV and JV-W acquired the data for the study. MN analyzed the data, together with IH, DB, MHV, and SH. MN and SH wrote the manuscript. All authors critically reviewed the manuscript.

FUNDING

MN, DB, and MHV are employees of SkylineDx. SH was funded by a fellowship of Abbott Diagnostics.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2021.793815/full#supplementary-material

REFERENCES

1. Vincent J-L, De Mendonça A, Cantraine F, Moreno R, Takala J, Suter PM, et al. Use of the SOFA score to assess the incidence of organ dysfunction/failure in intensive care units: results of a multicenter, prospective study. *Crit Care Med.* (1998) 26:1793–800. doi: 10.1097/00003246-199811000-00016
2. Le Gall J-R, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study. *JAMA.* (1993) 270:2957–63. doi: 10.1001/jama.1993.03510240069035
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II a severity of disease classification system. *Crit Care Med.* (1985) 13:818–29. doi: 10.1097/00003246-198510000-00009
4. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* (2020) 46:383–400. doi: 10.1007/s00134-019-05872-y
5. Shillan D, Sterne JAC, Champneys A, Gibbison B. Use of machine learning to analyse routinely collected intensive care unit data: a systematic review. *Crit Care.* (2019) 23:284. doi: 10.1186/s13054-019-2564-9
6. Song X, Alan SL, Kellum JA, Waitman LR, Matheny ME, Simpson SQ, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. *Nat Commun.* (2020) 11:5668. doi: 10.1038/s41467-020-19551-w
7. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R, et al. Review of challenges and opportunities in machine learning for health. *AMIA Summits Transl Sci Proc.* (2020) 2020:191. Available online at: https://arxiv.org/pdf/1806.00388.pdf
8. Henry KE, Hager DN, Pronovost PJ, Saria S. A targeted real-time early warning score (TREWScore) for septic shock. *Sci Transl Med.* (2015) 7:299ra122. doi: 10.1126/scitranslmed.aab3719
9. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation.* (2000) 101:E215–20. doi: 10.1161/01.cir.101.23.e215
10. Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Sci Data.* (2016) 3:160035. doi: 10.1038/sdata.2016.35
11. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models *via* coordinate descent. *J Stat Softw.* (2010) 33:1–22. doi: 10.18637/jss.v033.i01
12. Wunsch H, Angus DC, Harrison DA, Collange O, Fowler R, Hoste EAJ, et al. Variation in critical care services across North America and Western Europe. *Crit Care Med.* (2008) 36:2787–e8. doi: 10.1097/CCM.0b013e318186aec8
13. Wong A, Otles E, Donnelly JP, Krumm A, McCullough J, DeTroyer-Cooley O, et al. External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Intern Med.* (2021) 181:1065–70. doi: 10.1001/jamainternmed.2021.2626
14. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *BMJ.* (2020) 370:m3164. doi: 10.1136/bmj.m3164
15. Rivera SC, Liu X, Chan A-W, Denniston AK, Calvert MJ. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *BMJ.* (2020) 370:m3210. doi: 10.1136/bmj.m3210

16. Mandel JC, Kreda DA, Mandl KD, Kohane IS, Ramoni RB. SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J Am Med Informatics Assoc.* (2016) 23:899–908. doi: 10.1093/jamia/ocv189
17. Donnelly K. SNOMED-CT: the advanced terminology and coding system for eHealth. *Stud Health Technol Inform.* (2006) 121:279. Available online at: https://books.google.nl/books?hl=en&lr=&id=0vHjxr2G0CEC&oi=fnd&pg=PA279&dq=the+advanced+terminology+and+coding+system+for+eHealth&ots=c2qI4yBcmT&sig=FltqEsgx48qESeIjEs2EMEm6wkg&redir_esc=y#v=onepage&q=the%20advanced%20terminology%20and%20coding%20system%20for%20eHealth&f=false
18. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) the TRIPOD statement. *Circulation.* (2015) 131:211–9. doi: 10.1161/CIRCULATIONAHA.114.014508

Conflict of Interest: MN was employed by SkylineDx, Rotterdam and received a PhD fellowship from SkylineDx, Rotterdam. DB was employed by SkylineDx, Rotterdam. MVi was employed by SkylineDx, Rotterdam. SH received a fellowship from Abbott Diagnostics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Medicine

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fmed.2022.799811

A Novel Strategy for Predicting 72-h Mortality After Admission in Patients With Polytrauma: A Study on the Development and Validation of a Web-Based Calculator

Song Chen[1†], *Meiyun Liu*[2†], *Di Feng*[2], *Xin Lv*[2*] *and Juan Wei*[2*]

[1] *Department of Orthopaedic Trauma, East Hospital, School of Medicine, Tongji University, Shanghai, China,* [2] *Department of Anesthesiology, Shanghai Pulmonary Hospital, School of Medicine, Tongji University, Shanghai, China*

OPEN ACCESS

Edited by:
Longxiang Su,
Peking Union Medical College Hospital (CAMS), China

Reviewed by:
Adel Hamed Elbaih,
Suez Canal University, Egypt
Zhongheng Zhang,
Sir Run Run Shaw Hospital, China

***Correspondence:**
Juan Wei
sudaweijuan@163.com
Xin Lv
xinlvg@126.com

[†] *These authors have contributed equally to this work*

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *22 October 2021*
Accepted: *15 March 2022*
Published: *14 April 2022*

Citation:
Chen S, Liu M, Feng D, Lv X and Wei J (2022) A Novel Strategy for Predicting 72-h Mortality After Admission in Patients With Polytrauma: A Study on the Development and Validation of a Web-Based Calculator. Front. Med. 9:799811. doi: 10.3389/fmed.2022.799811

Background: Early and accessible screening of patients with polytrauma at a high risk of hospital death is essential. The purpose of this research was to seek an accurate and convenient solution to predict deaths occurring within 72 h after admission of these patients.

Methods: A secondary analysis was conducted on 3,075 patients with polytrauma from the Dryad database. We imputed missing values in eligible individuals with the k-nearest neighbor algorithm and then randomly stratified them into the training group (n = 2,461) and the validation group (n = 614) based on a proportion of 8:2. The restricted cubic spline, univariate, backward stepwise, and multivariate logistic regression methods were employed to determine the suitable predictors. Calibration and receiver operating characteristic (ROC) curves were applied to assess the calibration and discrimination of the obtained model. The decision curve analysis was then chosen as the measure to examine the clinical usage.

Results: Age, the Glasgow Coma Scale score, the Injury Severity Score, base excess, and the initial lactate level were inferred as independent prognostic factors related to mortality. These factors were then integrated and applied to construct a model. The performance of calibration plots, ROC curves, and decision curve analysis indicated that the model had satisfactory predictive power for 72-h mortality after admission of patients with polytrauma. Moreover, we developed a nomogram for visualization and a web-based calculator for convenient application (https://songandwen.shinyapps.io/DynNomapp/).

Conclusions: A convenient web-based calculator was constructed to robustly estimate the risk of death in patients with polytrauma within 72 h after admission, which may aid in further rationalization of clinical decision-making and accurate individual treatment.

Keywords: injury severity score, Glasgow Coma Scale, base excess, lactate, polytrauma, mortality, nomogram

INTRODUCTION

Trauma is the leading cause of death and disability in the world. More than 5 million deaths annually are due to injuries from falls, traffic accidents, landslides, and explosions, among others. Patients with polytrauma are the main contributors to this figure, accounting for 65 to 72% of the cases (1, 2). These patients are often severely injured, which is associated with hemorrhagic or traumatic shock and immune dysfunction, requiring accurate assessment and rapid treatment. Moreover, early screening of patients at risk of in-hospital death is crucial for ensuring patient safety, allocating medical resources appropriately, and reducing healthcare costs (3).

At present, various trauma scoring systems and hematological tests are suitable for evaluating the overall prognosis of patients with multiple traumas, and the introduction of internal environmental indicators, such as initial blood lactate, base excess (BE), and pH, in particular, provides early predictive evaluations for clinical purposes (4–6). However, these independent assessment methods are tedious to calculate, have too many scoring criteria, and have limited predictive power in assessing patient prognosis in the initial phase of trauma. Development of a simple and easy-to-use predictive model that incorporates factors related to the high risk of early death in patients with polytrauma is desirable.

Of all the models available, the logistic regression approach can provide a personalized, evidence-based, highly precise risk estimation in classification tasks. In addition, the advent of nomograms and network calculators has made the models user-friendly for disease prognosis and prediction, which facilitates decision-making related to patient management (7–9). Inspired by these efforts, this study aimed to develop and validate a prediction model and wrap it into a web-based calculator that allows rapid and precise individualized prediction of the risk of death within 72 h in patients with multiple traumas, by incorporating a few easily accessible clinical predictors.

METHODS

Data Source

The data sets yielded and analyzed are available from the Dryad Digital Repository, [https://datadryad.org/stash/dataset/doi.10.5061/dryad.bnzs7h45v]. The Dryad, an open resource database, provides a broad range of discoverable, freely reusable, and referable research data. Private information in the database has been anonymized. Data collection respects the principles outlined in the Declaration of Helsinki and has been approved by the local ethics committee.

Abbreviations: KNN, k-nearest neighbor; BMI, body mass index; ISS, Injury Severity Score; GCS, Glasgow Coma Scale; BE, base excess; IQR, interquartile range; OR, odds ratio; aOR, adjusted odds ratio; CI, confidence interval; ROC, receiver operator characteristic; AUC, area under the ROC curve; RCS, restricted cubic spline; VIF, variance inflation factor; DCA, decision curve analysis; PPV, positive predictive value; NPV, negative predictive value.

TABLE 1 | Baseline characteristics of the training and validation sets.

Characteristics	Training set ($N = 2461$)	Validation set ($N = 614$)	*P* value*
Age, years, median (IQR)	43 (28, 61)	43 (28, 61)	0.917
Sex, n (%)			
Female	642 (26.1)	170 (27.7)	0.421
Male	1,819 (73.9)	444 (72.3)	
BMI, kg/m^2,median (IQR)	24.7 (23.4, 26.1)	24.7 (23.4, 26.2)	0.687
ISS, median (IQR)	29 (22, 38)	27 (22, 36)	0.153
GCS, median (IQR)	6 (3, 14)	10 (3, 15)	0.225
pH, median (IQR)	7.34 (7.28, 7.38)	7.34 (7.29, 7.38)	0.400
BE, mmol/L, median (IQR)	−2.90 (−5.45, −1.10)	−2.70 (−5.30, −1.26)	0.632
Lactate, mmol/L, median (IQR)	2.30 (1.50, 3.50)	2.24 (1.50, 3.30)	0.195

**P-values between groups were assessed by chi-square and Mann-Whitney tests.*
BMI, body mass index; ISS, injury severity score; GCS, Glasgow Coma Scale; BE, base excess; IQR, interquartile range.

Study Design and Participants

A secondary retrospective analysis was performed based on the cohort study (10), which included multi-injury adult patients (>18 years old) treated at a Level I trauma center of the University Hospital Zurich from January 1, 1996 to January 1, 2013 and which excluded those with chronic diseases, oncological diseases, or genetic disorders that affect the musculoskeletal system. Time from injury to admission was defined as <24 h. Patients with multiple traumas were identified using an Injury Severity Score (ISS) of 16 or above, along with the criteria of the Berlin definition (11). Items selected from the data set for analysis are summarized in **Table 1**. The outcome was determined as patient's death within 72 h after admission. Related measurements of variables have been described carefully in the original article (10). Finally, among 3,668 patients recorded, 3,075 were enrolled, except for 579 (15.8%) with ISS values <16, 13 (0.4%) with no outcome data, and 1 (0.03%) with an incorrect body mass index (BMI) value marked as 0.

Missing Data

To maximize statistical power and minimize bias, k-nearest neighbor (KNN) (12) imputation with k equal to 10 was used to impute missing values in eligible patients. Then, the obtained imputation data were randomly stratified into two parts (i.e., training and validation cohorts) under a ratio of 8:2. We also carried out repeated analyses in the cohorts with complete data (i.e., data with all missing values removed) for comparison. Details on the statistical results are given in the **Supplementary Material**.

Sample Size Calculation

The *"pmsampsize"* package of R, version 4.0.2 (http://www.r-project.org/), was utilized to calculate the minimum training sample size required. Eight candidate predictor parameters were chosen to construct a multivariable prediction model for the binary outcome. Moreover, based on previous evidence (10), outcome prevalence is anticipated to be 0.268 (26.8%), and a lower bound for the new model's R-squared value is 0.288. For

the validation of sample size, a power calculation was carried out using PASS 15 (NCSS, LLC., Kaysville, UT, United States). The area under the receiver operating characteristic (ROC) curve (AUC; equivalent to the concordance statistic [C statistic]) was expected to be at least 0.8, and a two-tailed test with an alpha error of 0.05, beta error of 0.1, and power of 0.9 was conducted. As a result, the minimum sample size required for the training cohort is 302 patients with 81 events, while the validation cohort is 45 patients with 12 events. The eligible population is sufficient for model development and validation.

Statistical Analysis

Continuous variables were expressed as medians with interquartile ranges (IQRs) and were compared by unpaired Mann-Whitney test. Categorical variables were compared by χ^2 test. For each continuous variable at a significant level in the training cohort, we used a restricted cubic spline (RCS) with five knots at the 5, 35, 50, 65, and 95th percentiles to flexibly model its relationship with 72-h mortality after admission. Potential nonlinearity was tested using a likelihood ratio test comparing the model with only a linear term against the model with linear and cubic spline terms. Aiming to relax linear relationship assumptions, identified nonlinear continuous predictors were further categorized according to corresponding reference points determined by RCSs and horizontal lines with an odds ratio equal to 1. Then, linear continuous and acquired categorical predictors were examined with a univariate logistic regression approach for investigating the independent risk factors of mortality. All significant variables associated with death risks were candidates for stepwise multivariate analysis. To visualize the obtained model, a nomogram was generated according to multivariate logistic regression analysis outcomes and by applying the *"rms"* package. The predictive performance of the final model was measured by C statistic (13) and calibrated with 1,000 bootstrap samples for reducing overfitting bias. We also calculated the variance inflation factor (VIF) to examine the collinearity of each predictor in the prediction model and performed a formal sensitivity analysis, as described by Vander Weele and Ding (14), to capture the potential effect of unmeasured predictors on an obtained estimate.

For clinical utilization of the model, the total score for each patient was calculated from the nomogram. An ROC curve analysis was conducted to find optimal cutoff values that were determined by maximizing the Youden index (i.e., sensitivity + specificity−1). The accuracy of the optimal cutoff value was evaluated by the sensitivity, specificity, predictive values, and likelihood ratios. In addition, we established ROC curves for every predictor from the model. Pairwise comparisons of AUCs were tested with Delong's method. As a complement, decision curve analysis (DCA) was performed to quantify the clinical applicability of the model.

All the statistical analyses were completed with the R software. The remaining packages of R used were as follows: "*car*," "*caret*," "*splines*," "*pROC*," "*EValue*," "*rmda*," and "*ggplot2*".

A two-tailed test was carried out to determine the level of statistical difference, and $p < 0.05$ was considered statistically significant except in pairwise comparison of AUCs. In this scenario, *p*-values were adjusted by Bonferroni correction and tested with a bound of 0.003.

TABLE 2 | Baseline characteristics of patients who died or survived in the training cohort.

Characteristics	Alive (*N* = 1,911)	Dead (*N* = 550)	*P* value*
Age, years, median (IQR)	42 (27, 58)	51 (31, 72)	<0.001
Sex, *n* (%)			
Female	484 (25.3)	158 (28.7)	0.110
Male	1,427 (74.7)	392 (71.3)	
BMI, kg/m^2, median (IQR)	24.6 (23.1, 26.1)	24.9 (24.0, 26.0)	<0.001
ISS, median (IQR)	27 (21, 34)	34 (25, 50)	<0.001
GCS, median (IQR)	12 (3, 15)	3 (3, 3)	<0.001
pH, median (IQR)	7.35 (7.30, 7.38)	7.28 (7.19, 7.35)	<0.001
BE, mmol/L, median (IQR)	−2.55 (−4.40, −0.90)	−5.39 (−9.70, −2.40)	<0.001
Lactate, mmol/L, median (IQR)	2.10 (1.40, 3.04)	3.30 (2.20, 5.50)	<0.001

**P-values between groups were assessed by chi-square and Mann-Whitney tests.*
BMI, body mass index; ISS, injury severity score; GCS, Glasgow Coma Scale; BE, base excess; IQR, interquartile range.

RESULTS

Baseline Characteristics

A total of 3,075 adult patients with polytrauma were entered in the design data set. To account for missing data, KNN imputation was performed for BMI in 1,501 (48.8%), the Glasgow Coma Scale (GCS) score in 43 (1.4%), pH in 832 (27.1%), base excess (BE) in 703 (22.9%), and lactate in 472 (15.3%). The median patient age was 43 (IQR 28–61) years. In total, 2,263 (73.6%) patients were men and 687 (22.3%) died within 72 h after admission. A similar population distribution was detected in the complete data, except for the mortality rate of 117 (11%). This discrepancy may be due to the removal of a large amount of missing information, resulting in a biased estimate (**Supplementary Table S1**).

Among the 3,075 patients, 2,461 and 614 were assigned to the training and validation groups, respectively. Baseline characteristic distributions were similar between the cohorts. Mortality was 550 (22.3%) and 137 (22.3%) patients in the 2 groups, respectively (**Table 1**).

Compared to survivors in the training cohort, those who died showed a higher rate of age (42 [IQR 27–58] vs. 51 [IQR 31–72], $P < 0.001$), BMI (24.6 [IQR 23.1–26.1] vs. 24.9 [IQR 24–26], $P < 0.001$), ISS (27 [IQR 21–34] vs. 34 [IQR 25–50], $P < 0.001$), and lactate (2.1 [IQR 1.4–3.04] vs. 3.3 [IQR 2.2–5.5], $P < 0.001$) and presented a lower value in the GCS score (12 [IQR 3–15] vs. 3 [IQR 3–3], $P < 0.001$), pH (7.35 [IQR 7.3–7.38] vs. 7.28 [IQR 7.19–7.35], $P < 0.001$), and BE (−2.55 [IQR −4.40–−0.9] vs. −5.39 [IQR −9.7–−2.4], $P < 0.001$). No statistical difference was detected in gender between the two cohorts (**Table 2**).

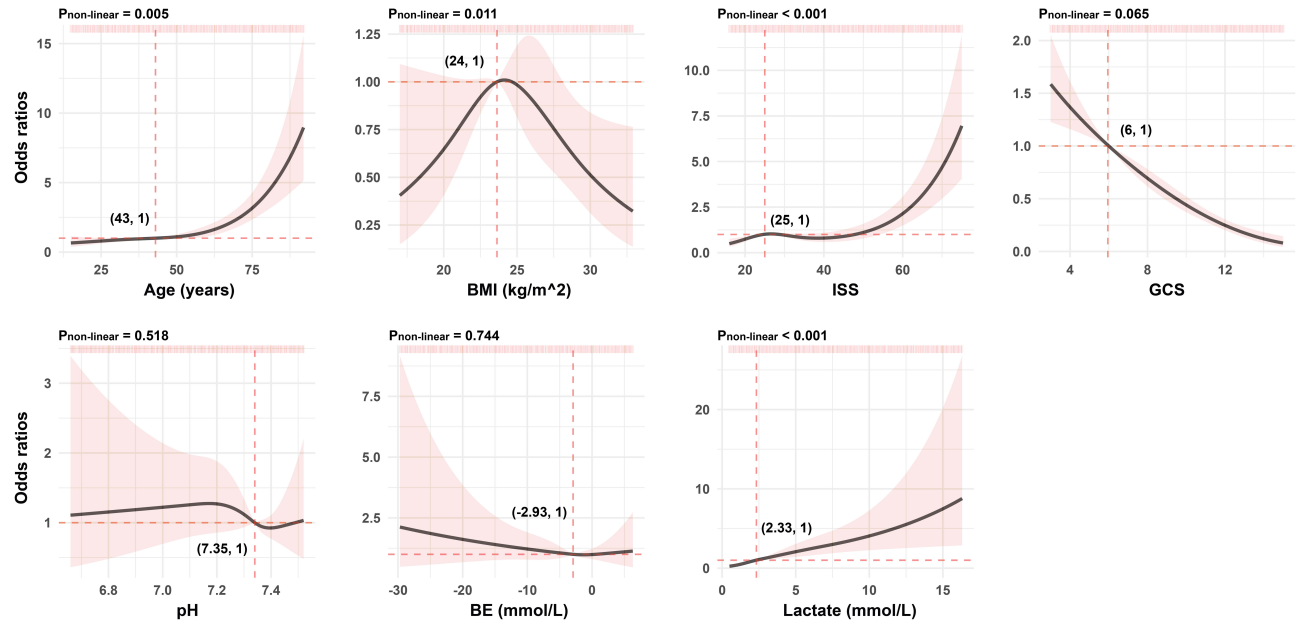


FIGURE 1 | Association between continuous predictors and 72-h mortality in patients with polytrauma by the RCS analysis. For each curve, five knots at the 5th, 35th, 50th, 65th, and 95th percentiles were selected. Solid lines represent odds ratios, and shaded areas represent 95% CIs. The reference point is the first value closest or equal to the odds ratio at 1 (i.e., the intersection of the two red dashed lines). RCS, restricted cubic spline; BMI, body mass index; ISS, injury severity score; GCS, Glasgow Coma Scale; BE, base excess.

TABLE 3 | Comparisons between patients who survived and deceased patients across all post-conversion variables in the training cohort.

Variables	Alive (*N* = 1911)	Dead (*N* = 550)	*P* value*
Age, years, *n* (%)			<0.001
<43	978 (51.2)	215 (39.1)	
≥43	933 (48.8)	335 (60.9)	
BMI, kg/m^2,n (%)			<0.001
< 24	717 (37.5)	129 (23.5)	
≥ 24	1,194 (62.5)	421 (76.5)	
ISS, *n* (%)			<0.001
< 25	679 (35.5)	58 (10.5)	
≥ 25	1232 (64.5)	492 (89.5)	
Lactate, mmol/L, *n* (%)			<0.001
<2.33	1,095 (57.3)	152 (27.6)	
≥2.33	816 (42.7)	398 (72.4)	

**P-values between groups were assessed by the chi-square test.*
BMI, body mass index; ISS, injury severity score.

Model Specifications and Predictors

As presented in **Figure 1**, continuous variables like age, BMI, ISS, and lactate do not meet the linear relationship assumptions (All $P_{\text{non-linear}} < 0.05$). We converted these variables to categorical variables with reference points as cutoff values for the next univariable logistic analysis. Comparisons between those who survived and deceased patients were significant across all post-conversion variables (**Table 3**).

TABLE 4 | A logistic regression analysis of the 72-h mortality for patients in the training cohort.

Variables	Univariable OR (95% CI)	*P* value	Multivariable aOR (95% CI)	*P* value
Factors selected by stepwise analysis				
Age, years				
<43	1 [Reference]	NA	1 [Reference]	NA
≥43	1.63 (1.35, 1.98)	<0.001	2.25 (1.78, 2.83)	< 0.001
ISS				
<25	1 [Reference]	NA	1 [Reference]	NA
≥25	4.66 (3.52, 6.28)	<0.001	2.96 (2.15, 4.07)	< 0.001
Lactate, mmol/L				
<2.33	1 [Reference]	NA	1 [Reference]	NA
≥2.33	3.51 (2.86, 4.33)	<0.001	2.16 (1.67, 2.81)	< 0.001
GCS	0.78 (0.76, 0.81)	<0.001	0.80 (0.78, 0.83)	< 0.001
pH	0.003 (0.001, 0.007)	<0.001	0.32 (0.08, 1.23)	0.097
BE, mmol/L	0.86 (0.84, 0.88)	<0.001	0.94 (0.91, 0.98)	0.001
Factors not selected by stepwise analysis				
BMI, kg/m^2				
<24	1 [Reference]	NA	NA	NA
≥24	1.96 (1.58, 2.44)	<0.001		

BMI, body mass index; ISS, injury severity score; GCS, Glasgow Coma Scale; BE, base excess; OR, odds ratio; aOR, adjusted odds ratio; CI, confidence interval; NA, not applicable.

The results of the univariate logistic analysis are shown in **Table 4**. Backward stepwise selection with AIC determined the following 6 variables that were most strongly associated with

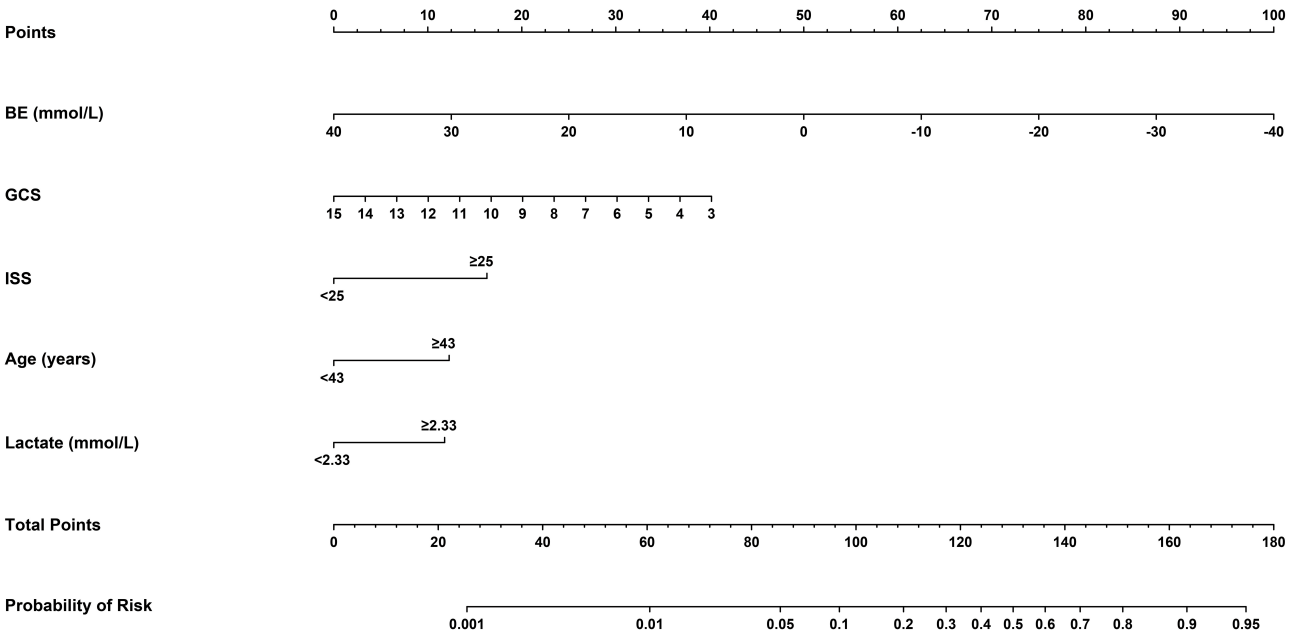


FIGURE 2 | Nomogram for estimating the risk of 72-h mortality after admission in patients with polytrauma. Using a nomogram, we first determined the location of the axis corresponding to each variable and drew a vertical line to the "Points" axis to get a score, then, summed scores from all variables, and drew another vertical line from the "Total Points" axis to the "Probability of Risk" axis for the predicted probability. BE, base excess; GCS, Glasgow Coma Scale; ISS, injury severity score.

death risk: age, ISS, lactate, GCS, pH, and BE. In the multivariable analysis, age of at least 43 years old (OR 2.25; 95% confidence interval [CI] 1.78–2.83; $P < 0.001$), ISS of at least 25 (OR 2.96; 95% CI 2.15–4.07; $P < 0.001$), lactate of at least 2.33 (OR 2.16; 95% CI 1.67–2.81; $P < 0.001$), GCS score (OR.8; 95% CI 0.78–0.83; $P < 0.001$), and BE (OR 0.94; 95% CI 0.91–0.98; $P = 0.001$) were all independently related to mortality (**Table 4**). Similar findings are obtained in the complete data, as shown in **Supplementary Table S3**.

Model Development and Validation

The identified independently associated risk factors were then applied to construct the final model and form a nomogram for estimating 72-h mortality risk after admission (**Figure 2**). For predictors from the model, VIFs were 2.76 or less, indicating the absence of collinearity. Moreover, the E-value, a standard way to quantify the potential effect of unmeasured predictors on the obtained estimate, for each predictor is calculated and presented in **Supplementary Figure S1**. The lowest one is 1.21; that is to say, our estimates were robust to unmeasured confounders, except in the case of a strong unmeasured confounder that was substantially associated with death risk. In order to simplify the clinical application of the model, we also designed a web-based calculator (https://songandwen.shinyapps.io/DynNomapp/) to predict death risks for patients with polytrauma (**Supplementary Figure S2**).

Regarding model performance testing, we first completed internal validation with the bootstrap method. The model had good discrimination in assessing mortality, with an unadjusted C statistic of 0.85 (95% CI 0.83–0.86) and a bootstrap-corrected C statistic of 0.85. Besides, the calibration plots indicated a good agreement between risk estimates and actual deaths (**Figure 3A**), whereas in the validation data set, the model presented a C statistic of 0.84 (95% CI 0.81–0.88) for predicting mortality. There was also a satisfactory calibration curve for risk estimation (**Figure 3B**).

To further compare the predictive value of the model, age, ISS, GCS, BE, and lactate for the 72-h mortality in patients with polytrauma, ROC curves were plotted (**Figure 4**). In the training cohort, relative AUCs were 0.85 (95% CI:0.83–0.86),0.6 (95% CI:0.57–0.63),0.67 (95% CI:0.66–0.69),0.76 (95% CI:0.74–0.77),0.7 (95% CI:0.67–0.73), and 0.71 (95% CI:0.69–0.74). There were no statistical differences among ISS, GCS, BE, and lactate (all $P > 0.003$), while differences between our model and any of the others were statistically significant (all $P < 0.001$) as well as age ($P < 0.001$). As for validation, the AUCs showed modest changes and were 0.84 (95% CI:0.81–0.88),0.62 (95% CI:0.56–0.68),0.69 (95% CI:0.64–0.74),0.76 (95% CI:0.73–0.8),0.69 (95% CI:0.64–0.75), and 0.69 (95% CI:0.64–0.74). Differences between the model and any of the others remained statistically significant (all $P < 0.001$), but no significance was observed in the remaining comparisons (all $P > 0.003$). The above evidence suggested that our model had superior predictive performance over any of the single predictors mentioned.

Clinical Usage of the Model

We assumed that a patient with a nomogram score above a defined threshold was at high death risk, but that with the

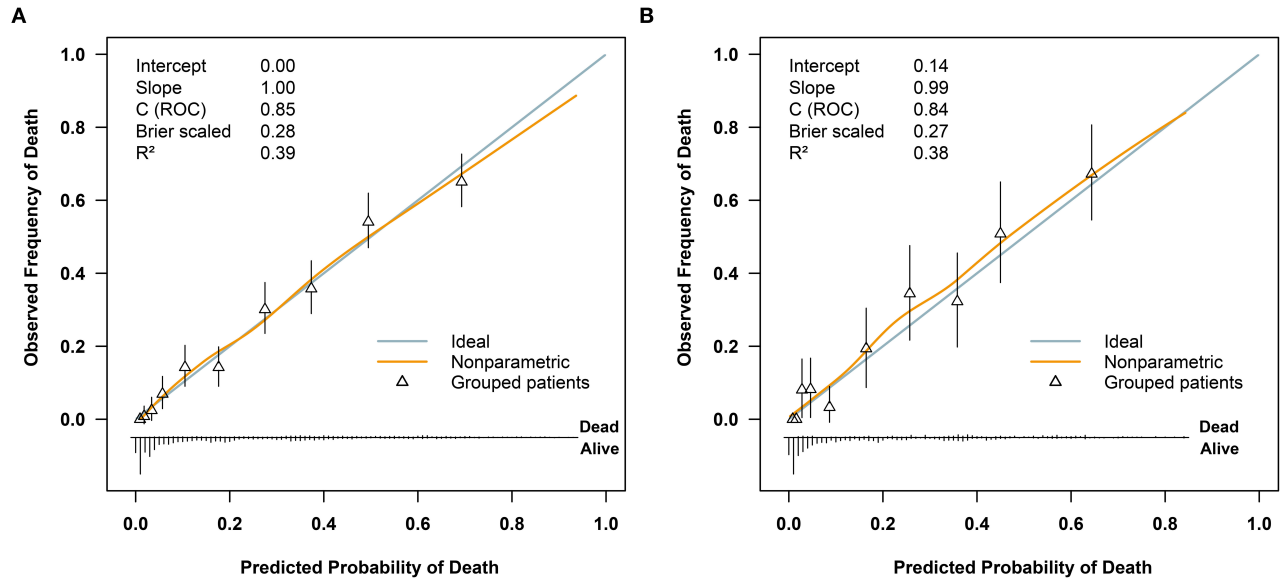


FIGURE 3 | Receiver operating characteristic curves for validating the discrimination of our model and included variables: **(A)** in the training cohort ($n = 2461$) and **(B)** in the validation cohort ($n = 614$). ISS, injury severity score; GCS, Glasgow Coma Scale; BE, base excess.

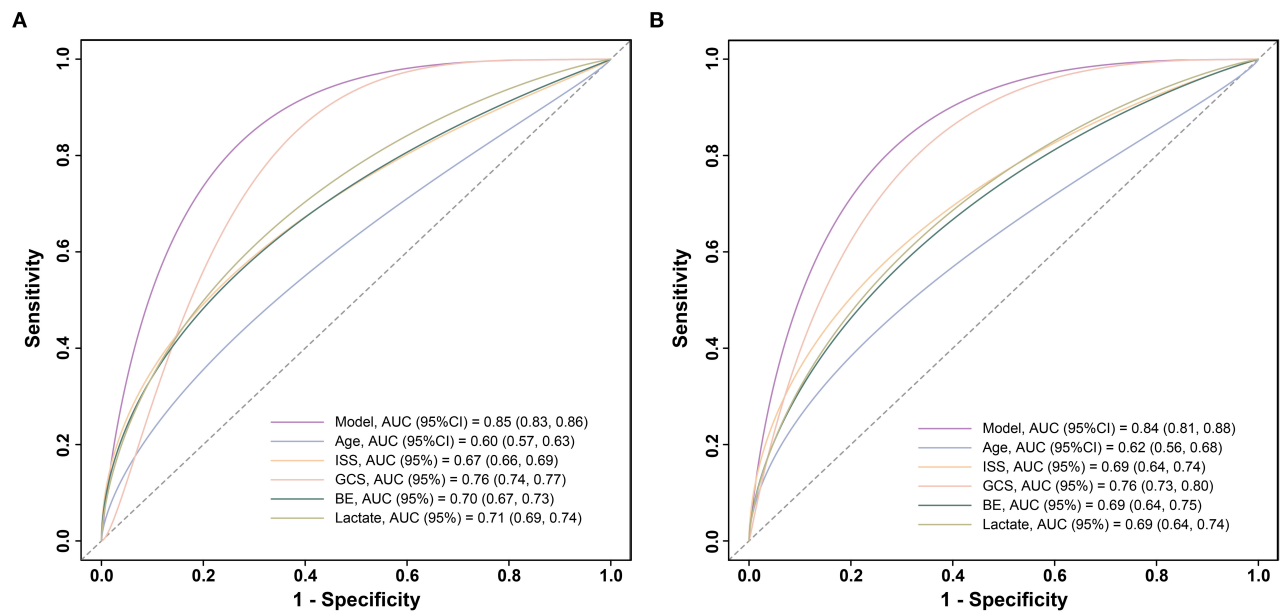


FIGURE 4 | Calibration plots of the model for predicting 72-h mortality in patients with polytrauma **(A)** in the training cohort ($n = 2461$) and **(B)** in the validation cohort ($n = 614$). ROC, receiver operator characteristic.

defined threshold was at low death risk. Then, a total score was calculated for each patient and an optimal cutoff value of 111 was determined. On this basis, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were 81.5, 74, 47.4, and 93.3% in the training cohort and 82.5, 73.8, 47.5, and 93.6% in the validation cohort, respectively (**Table 5**).

Furthermore, we utilized DCAs to evaluate the net benefit of the model for decision-making. As illustrated in **Figure 5**, in the training cohort, the model is applicable when the threshold is between 0.01 and 0.9, as net benefits are >0, while the validation cohort has a valid range of between 0.01 and 0.74.

DISCUSSION

It is vital to evaluate the clinical condition of patients with multiple traumas in an early stage. From the views of most authors, estimating early mortality plays an equally important role in predicting subsequent complications (15–18). In this study, age, ISS, GCS, BE, and lactate were identified as independent risk factors for early death in patients with polytrauma based on the information extracted from an online database. The proposed model, which incorporated the above 5 readily accessible variables, performed impressively, being backed by C statistic values of 0.85 and 0.84 in the training and validation cohorts, respectively, and calibration curves indicated an excellent agreement between predictions and actual observations. Furthermore, our results revealed that age, ISS, GCS, BE, and lactate were less accurate in predicting mortality than our new model. To adapt the model to clinical practice, we summarized sensitivity, specificity, NPV, and PPV for assessing mortality risk by considering 111 as the cutoff value (**Table 5**). Patients with scores of 111 or more (1,185 of 3,075, [38.5%]) are selected into the high-risk subgroup. The corresponding net benefit values are 0.19 in the training cohort and 0.2 in the validation cohort. The abovementioned evidence support that our model might serve as an efficient and reliable tool for estimating the risk of 72-h death after admission and might aid in clinical decision-making.

The Glasgow Coma Scale and the ISS are routine initial assessment scoring systems for patients with polytrauma. The former, an essential measurement of neurological function and severity of the head injury, has advantages such as being simple, practical, time-efficient, and cost-effective (19). Several authors have determined that low GCS was associated with poor outcome (20–22), which is consistent with our evidence from the RCS of GCS in **Figure 1**. The ISS is another major predictor of trauma mortality and focuses on anatomical scoring. Contrary to the GCS scores, the probability of patient survival decreases with increasing ISS scores (23). Watts et al. (24) reported that ISS was positively associated with in-hospital mortality in elderly patients with trauma. It is not unique. In our findings, compared to patients with ISS <25, the risk of death was approximately three-fold greater in patients with larger values. However, the above scoring systems fail to assess both physiological disorders and anatomical damages in patients with polytrauma. In addition, there is always confusion on how to scientifically

TABLE 5 | Performance metrics of the nomogram for estimating the risk of 72-h mortality after admission.

	Value (95% CI)	
Performance metrics	**Training set**	**Validation set**
Cutoff score*	111	111
Sensitivity, %	81.5 (77.9, 84.6)	82.5 (75.1, 88.4)
Specificity, %	74.0 (72.0, 75.9)	73.8 (69.6, 77.7)
Positive predictive value, %	47.4 (44.9, 53.0)	47.5 (42.4, 59.5)
Negative predictive value, %	93.3 (91.8, 93.9)	93.6 (90.4, 94.8)
Positive likelihood ratio	3.13 (2.88, 3.41)	3.15 (2.66, 3.73)
Negative likelihood ratio	0.25 (0.21, 0.30)	0.23 (0.16, 0.34)

**Optimal cutoff scores were determined by maximizing the Youden index (i.e., sensitivity + specificity−1). CI, confidence interval.*

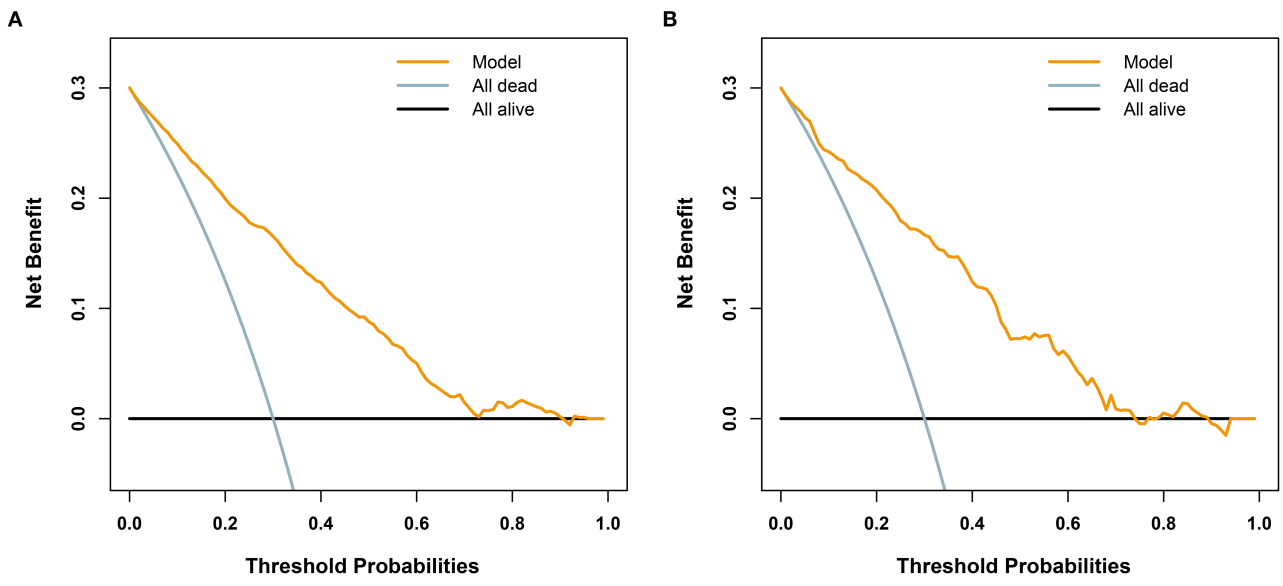


FIGURE 5 | Decision analysis curves for the developed model for predicting 72-h mortality in patients with polytrauma. The black line represents a scheme to make all patients alive. The light blue line represents an improper protocol, leading to all patient mortality that occurred during the course of the study. The yellow line represents net benefits of the clinical application of the model. **(A)** In the training cohort ($n = 2461$), when the threshold is between 0.01 and 0.9, the model is applicable, since the net benefit is >0. **(B)** In the validation cohort (n = 614), the valid range is between 0.01 and 0.74.

synthesize the results of multiple scoring systems to guide next clinical implementations. Fortunately, none of these are issues for the model-based network calculator, and all it takes is to correctly categorize ISS values and determine the corresponding GCS values.

Age is also a predictor for the risk of mortality related to multiple injuries but not a part of the above two scoring systems (25). In a meta-analysis of older adults with trauma, Hashmi et al. (26) found that the risk of death increased with age and was two times higher in patients aged 74 than in those aged 62. We agree with the efforts of Hashmi et al. In addition, we noticed that the risk kept increasing faster when the patient was over 50 years old (**Figure 1**). In comparison with those younger than 43, the older ones were exposed to an additional 2.25-fold risk of death (OR 2.25; 95% CI 1.78–2.83). This reminds us that more attention should be paid to both middle-aged and older patients with multiple injuries. The subtle gap between our results and those of Hashmi et al. (26) is possibly due to various observed populations and different statistical methods chosen for analysis. In any case, it is certainly a sensible move to include age in our model to improve the accuracy of the estimates.

Admission BE is a recognized trauma marker that can evaluate injury severity and forecast post-traumatic outcomes (27). Several studies have shown that an initial negative BE predicts the mortality risk of patients with trauma, meaning that the poorer the BE, the higher the in-hospital mortality (28–30). Across such research, we could observe a trend toward higher mean BE in survival than in death, which is also reflected in our study (−2.55 [IQR −4.40–−0.0] vs. −5.39 [IQR −9.7–−2.4], $P < 0.001$). Lichtveld et al. (28) concluded that BE was an independent predictor of mortality in patients with trauma, with an OR of 0.92 (95% CI:0.89–0.95), indicating an 8% increase in the risk of death for each unit reduction in BE. Our findings were in close agreement with those published by Lichtveld et al. The OR is 0.94 (95% CI:0.91–0.98), suggesting that every 1 mmol/L decrease in BE was associated with a 6% increase in the risk of death. These slight discrepancies may be explained by the different enrollment populations of the two studies. Furthermore, the effect of BE on 72-h mortality was also assessed by the ROC analysis. However, the measured AUCs were still significantly lower than the fit of the model we developed.

Lactate is a usual clinical biomarker for diagnosing shock and monitoring resuscitation. It is valuable not only for patients with sepsis shock (31) but also for patients with trauma. In a study including 1,829 patients with blunt trauma, Gale et al. (32) confirmed that the initial lactate was a dependable prognosticator of patients at a higher risk of in-hospital mortality. In another observational cohort study with 1,075 patients with trauma, Raux et al. (33) observed that admission lactate was superior in predicting early deaths, severe traumatic lesions, and massive hemorrhage. According to our research, patients with admission lactate over the bound of 2.33 mmol/L were at 2.16-fold risk of 72-h death than others (OR 2.16; 95% CI 1.67–2.81). The AUC of ROC used to estimate the predictive value of lactate for 72-h mortality was 0.71 in the training group and 0.69 in the validation group, which is close to 0.716 reported by Sammour (34) and lower than the performance of our model.

Moreover, we opted for a logistic regression approach to construct the prediction model, which may be limited by its linearity assumption. Although great care has been taken to build RCS models for exploring this assumption, the residual predictor-response variable complex relationship may still have been overlooked. These challenges may be easily solved by machine learning algorithms, which do not require assumptions of strict data structure and have the ability to learn complex functional forms with nonparametric methods. Furthermore, ensemble modeling methods, i.e., combining two or more machine learning algorithms, can be applied for the improvement of prediction accuracy (35). We are considering the application of this promising technology as an alternative in future research.

Limitations

Our study has several limitations. First, in the final prediction model, missing data were present for the BMI, GCS, pH, BE, and lactate variables. The data were considered to be missing at random, and KNN imputation was conducted to minimize selection bias. We repeated the analysis on the complete data and obtained similar results. Second, this is a secondary analysis with fixed data, and we may have overlooked some important predictors. Therefore, we quantified the unmeasured factors to assess the robustness of our model. Third, all the analyses were conducted on the basis of data from one institution; there is a need to validate the results from other centers. Further prospective studies are also required to affirm the dependability of our model. Hence, the web calculator was developed to make these requirements easy to implement. In addition, although the model reached good predictive accuracy with a cutoff of 111, the false positive and false negative rates were 26 and 18.5% in the training cohort and 26.2 and 17.5% in the validation cohort, respectively. For 72-h mortality prediction, the performance of our model still needs to be improved to make meaningful clinical decisions.

CONCLUSIONS

A clinical prediction model was constructed and wrapped into a web-based calculator to estimate mortality risk easily and robustly in patients with polytrauma within 72 h of hospital admission, which may contribute to further rationalization of clinical decision-making and accurate individual treatment. Under another aspect, the calculator may identify patients at a high risk of death and thus avoid corresponding adverse events.

DATA AVAILABILITY STATEMENT

The data utilized and analyzed in the current study is available in the Dryad database. The website is https://doi.org/10.5061/dryad.bnzs7h45v.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committees of the East Hospital of Tongji University, Shanghai, China. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

SC, ML, DF, and JW designed the study. SC, ML, and JW performed data analysis and prepared the manuscript. JW and XL contributed to improving the article and funding for the project. All authors reviewed the manuscript, read, and approved the submitted manuscript.

FUNDING

This study was supported by grants from the Clinical Research Plan of SHDC (No. SHDC2020CR3044B) and the Clinical Research Special Project of Shanghai Municipal Health Commission (No. 202040004).

ACKNOWLEDGMENTS

We are very grateful to Halvachizadeh et al. for their data collection and sharing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2022.799811/full#supplementary-material

REFERENCES

1. Breugel JMM, Niemeyer MJS, Houwert RM, Groenwold RHH, Leenen LPH, Wessem KJP. Global changes in mortality rates in polytrauma patients admitted to the ICU-a systematic review. *World J Emerg Surg.* (2020) 15:55. doi: 10.1186/s13017-020-00330-3
2. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* (2018) 392:1789–858. doi: 10.1016/S0140-6736(18)32279-7
3. Negrin LL, Antoni A, Hajdu S, Heinz T. A Novel Approach to Identify Polytraumatized Patients in Extremis. *Biomed Res Int.* (2018) 2018:73201–58. doi: 10.1155/2018/7320158
4. Qi J, Bao L, Yang P, Chen D. Comparison of base excess, lactate and pH predicting 72-h mortality of multiple trauma. *BMC Emerg Med.* (2021) 21:80. doi: 10.1186/s12873-021-00465-9
5. Vogel JA, Newgard CD, Holmes JF, Diercks DB, Arens AM, Boatright DH, et al. Validation of the Denver Emergency Department Trauma Organ Failure Score to Predict Post-Injury Multiple Organ Failure. *J Am Coll Surg.* (2016) 222:73–82. doi: 10.1016/j.jamcollsurg.2015.10.010
6. Wardi G, Brice J, Correia M, Liu D, Self M, Tainter C. Demystifying Lactate in the Emergency Department. *Ann Emerg Med.* (2020) 75:287–98. doi: 10.1016/j.annemergmed.2019.06.027
7. Graesslin O, Abdulkarim BS, Coutant C, Huguet F, Gabos Z, Hsu L, et al. Nomogram to predict subsequent brain metastasis in patients with metastatic breast cancer. *J Clin Oncol.* (2010) 28:2032–7. doi: 10.1200/JCO.2009.24.6314
8. Kattan MW, Scardino PT. Evidence for the usefulness of nomograms. *Nat Clin Pract Urol.* (2007) 4:638–9. doi: 10.1038/ncpuro0968
9. Wang J, He L, Tang Y, Li D, Yang Y, Zeng Z. Development and validation of a nomogram with an epigenetic signature for predicting survival in patients with lung adenocarcinoma. *Aging (Albany NY).* (2020) 12:23200–16. doi: 10.18632/aging.104090
10. Halvachizadeh S, Baradaran L, Cinelli P, Pfeifer R, Sprengel K, Pape HC. How to detect a polytrauma patient at risk of complications: A validation and database analysis of four published scales. *PLoS ONE.* (2020) 15:e0228082. doi: 10.1371/journal.pone.0228082
11. Pape HC, Lefering R, Butcher N, Peitzman A, Leenen L, Marzi I, et al. The definition of polytrauma revisited: An international consensus process and proposal of the new 'Berlin definition'. *J Trauma Acute Care Surg.* (2014) 77:780–6. doi: 10.1097/TA.0000000000000453
12. Altman NS. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am Stat.* (1992) 46:175–85. doi: 10.1080/00031305.1992.10475879
13. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA.* (1982) 247:2543–6. doi: 10.1001/jama.1982.03320430047030
14. VanderWeele TJ, Ding P. Sensitivity Analysis in Observational Research: Introducing the E-Value. *Ann Intern Med.* (2017) 167:268–74. doi: 10.7326/M16-2607
15. Domingues CA, Coimbra R, Poggetti RS, Nogueira LS, de Sousa RMC. New Trauma and Injury Severity Score (TRISS) adjustments for survival prediction. *World J Emerg Surg.* (2018) 13:12. doi: 10.1186/s13017-018-0171-8
16. Eichelberger MR, Bowman LM, Sacco WJ, Mangubat EA, Lowenstein AD, Gotschall CS. Trauma score versus revised trauma score in TRISS to predict outcome in children with blunt trauma. *Ann Emerg Med.* (1989) 18:939–42. doi: 10.1016/S0196-0644(89)80457-3
17. Napolitano LM, Fulda GJ, Davis KA, Ashley DW, Friese R, Van Way CW 3rd, et al. Challenging issues in surgical critical care, trauma, and acute care surgery: a report from the Critical Care Committee of the American Association for the Surgery of Trauma. *J Trauma.* (2010) 69:1619–33. doi: 10.1097/TA.0b013e3182011089
18. Regel G, Lobenhoffer P, Grotz M, Pape HC, Lehmann U, Tscherne H. Treatment results of patients with multiple trauma: an analysis of 3406 cases treated between 1972 and 1991 at a German Level I Trauma Center. *J Trauma.* (1995) 38:70–8. doi: 10.1097/00005373-199501000-00020
19. Davis DP, Vadeboncoeur TF, Ochs M, Poste JC, Vilke GM, Hoyt DB. The association between field Glasgow Coma Scale score and outcome in patients undergoing paramedic rapid sequence intubation. *J Emerg Med.* (2005) 29:391–7. doi: 10.1016/j.jemermed.2005.04.012
20. Lieberman JD, Pasquale MD, Garcia R, Cipolle MD, Mark Li P, Wasser TE. Use of admission Glasgow Coma Score, pupil size, and pupil reactivity to determine outcome for trauma patients. *J Trauma.* (2003) 55:437–42; discussion 442-3. doi: 10.1097/01.TA.0000081882.79587.17
21. Mauritz W, Leitgeb J, Wilbacher I, Majdan M, Janciak I, Brazinova A, et al. Outcome of brain trauma patients who have a Glasgow Coma Scale score of 3 and bilateral fixed and dilated pupils in the field. *Eur J Emerg Med.* (2009) 16:153–8. doi: 10.1097/MEJ.0b013e32832a0864
22. Tien HC, Cunha JR, Wu SN, Chughtai T, Tremblay LN, Brenneman FD, et al. Do trauma patients with a Glasgow Coma Scale score of 3 and bilateral fixed and dilated pupils have any chance of survival? *J Trauma.* (2006) 60:274–8. doi: 10.1097/01.ta.0000197177.13379.f4
23. Elgin LB, Appel SJ, Grisham D, Dunlap S. Comparisons of Trauma Outcomes and Injury Severity Score. *J Trauma Nurs.* (2019) 26:199–207. doi: 10.1097/JTN.0000000000000449
24. Watts HF, Kerem Y, Kulstad EB. Evaluation of the revised trauma and injury severity scores in elderly trauma patients. *J Emerg Trauma Shock.* (2012) 5:131–4. doi: 10.4103/0974-2700.96481

25. Bruijns SR, Guly HR, Bouamra O, Lecky F, Lee WA. The value of traditional vital signs, shock index, and age-based markers in predicting trauma mortality. *J Trauma Acute Care Surg.* (2013) 74:1432–7. doi: 10.1097/TA.0b013e31829246c7
26. Hashmi A, Ibrahim-Zada I, Rhee P, Aziz H, Fain MJ, Friese RS, et al. Predictors of mortality in geriatric trauma patients: a systematic review and meta-analysis. *J Trauma Acute Care Surg.* (2014) 76:894–901. doi: 10.1097/TA.0b013e3182ab0763
27. Ibrahim I, Chor WP, Chue KM, Tan CS, Tan HL, Siddiqui FJ, et al. Is arterial base deficit still a useful prognostic marker in trauma? A systematic review. *Am J Emerg Med.* (2016) 34:626–35. doi: 10.1016/j.ajem.2015.12.012
28. Lichtveld RA, Panhuizen IF, Smit RB, Holtslag HR, van der Werken C. Predictors of Death in Trauma Patients who are Alive on Arrival at Hospital. *Eur J Trauma Emerg Surg.* (2007) 33:46–51. doi: 10.1007/s00068-007-6097-6
29. Wang SY, Liao CH, Fu CY, Kang SC, Ouyang CH, Kuo IM, et al. An outcome prediction model for exsanguinating patients with blunt abdominal trauma after damage control laparotomy: a retrospective study. *BMC Surg.* (2014) 14:24. doi: 10.1186/1471-2482-14-24
30. Soderlund T, Ikonen A, Pyhalto T, Handolin L. Factors associated with in-hospital outcomes in 594 consecutive patients suffering from severe blunt chest trauma. *Scand J Surg.* (2015) 104:115–20. doi: 10.1177/1457496914543976
31. Dettmer M, Holthaus CV, Fuller BM. The impact of serial lactate monitoring on emergency department resuscitation interventions and clinical outcomes in severe sepsis and septic shock: an observational cohort study. *Shock.* (2015) 43:55–61. doi: 10.1097/SHK.0000000000000260
32. Gale SC, Kocik JF, Creath R, Crystal JS, Dombrovskiy VY. A comparison of initial lactate and initial base deficit as predictors of mortality after severe blunt trauma. *J Surg Res.* (2016) 205:446–55. doi: 10.1016/j.jss.2016.06.103
33. Raux M, Le Manach Y, Gauss T, Baumgarten R, Hamada S, Harrois A, et al. Comparison of the Prognostic Significance of Initial Blood Lactate and Base Deficit in Trauma Patients. *Anesthesiology.* (2017) 126:522–33. doi: 10.1097/ALN.0000000000001490
34. Sammour T, Kahokehr A, Caldwell S, Hill AG. Venous glucose and arterial lactate as biochemical predictors of mortality in clinically severely injured trauma patients–a comparison with ISS and TRISS. *Injury.* (2009) 40:104–8. doi: 10.1016/j.injury.2008.07.032
35. Zhang Z, Chen L, Xu P, Hong Y. Predictive analytics with ensemble modeling in laparoscopic surgery: a technical note. *Laparosc Endosc Robot Surg.* (2022). doi: 10.1016/j.lers.2021.12.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Medicine

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fmed.2022.766447

Establishment and Implementation of Potential Fluid Therapy Balance Strategies for ICU Sepsis Patients Based on Reinforcement Learning

Longxiang Su[1†], Yansheng Li[2†], Shengjun Liu[1†], Siqi Zhang[2], Xiang Zhou[1], Li Weng[3], Mingliang Su[2], Bin Du[3], Weiguo Zhu[4*] and Yun Long[1*]*

[1] Department of Critical Care Medicine, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China, [2] DHC Mediway Technology Co., Ltd., Beijing, China, [3] Medical Intensive Care Unit, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [4] Department of Information Center, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Yue Dong,
Mayo Clinic, United States

Reviewed by:
Huiqing Ge,
Sir Run Run Shaw Hospital, China
Bo Hu,
Zhongnan Hospital of Wuhan University, China

***Correspondence:**
Bin Du
dubin98@gmail.com
Weiguo Zhu
zhuwg@pumch.cn
Yun Long
icu_longyun@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Intensive Care Medicine and Anesthesiology, a section of the journal Frontiers in Medicine

Received: *29 August 2021*
Accepted: *17 February 2022*
Published: *14 April 2022*

Citation:
Su L, Li Y, Liu S, Zhang S, Zhou X, Weng L, Su M, Du B, Zhu W and Long Y (2022) Establishment and Implementation of Potential Fluid Therapy Balance Strategies for ICU Sepsis Patients Based on Reinforcement Learning. Front. Med. 9:766447. doi: 10.3389/fmed.2022.766447

Objective: Fluid therapy for sepsis patients has always been a problem that puzzles clinicians, that is, knowing when patients need fluid infusion and when they need negative fluid balance. Different clinicians may have different judgment criteria and make different decisions. Recently, studies have suggested that different fluid treatment strategies can cause different clinical outcomes. This study is intended to establish and verify a model for judging the direction of fluid therapy based on machine learning.

Method: This study included 2705 sepsis patients from the Peking Union Medical College Hospital Intensive Care Medical Information System and Database (PICMISD) from January 2016 to April 2020. The training set and test set (January 2016 to June 2019) were randomly divided. Twenty-seven features were extracted for modeling, including 25 state features (bloc, vital sign, laboratory examination, blood gas assay and demographics), 1 action feature (fluid balance) and 1 outcome feature (ICU survival or death). SARSA was used to learn the data rules of the training set. Deep Q-learning (DQN) was used to learn the relationship between states and actions of the training set and predict the next balance. A double-robust estimator was used to evaluate the average expected reward of the test set in the deep Q-learning model. Lastly, we verified the difference between the predicted fluid therapy model and the actual treatment for the patient's prognoses, with sepsis patient data from July 2019 to April 2020 as the validation set.

Results: The training set and test set were extracted from the same database, and the distribution of liquid balance was similar. Actions were divided into five intervals corresponding to 0–20, 20–40, 40–60, 60–80, and 80–100% percentiles of fluid balance. The higher the reward of $Q(s, a)$ calculated by SARSA from the training set, the lower the mortality rate. Deep Q-learning indicates that both fluid balance differences that are too high and too low show an increase in mortality. The more consistent the fluid balance prediction with the real result, the lower the mortality rate. The smaller the difference between the prediction and the reality, the lower the mortality rate. The double-robust

estimator shows that the model has satisfactory stability. The validation set indicates that the mortality rate of patients in the "predicted negative fluid balance and actual negative fluid balance" subgroup was the lowest, which was statistically significant, indicating that the model can be used for clinical verification.

Conclusion: We used reinforcement learning to propose a possible prediction model for guiding the direction of fluid therapy for sepsis patients in the ICU. This model may accurately predict the best direction for fluid therapy, thereby improving patient prognosis.

Keywords: sepsis, fluid therapy, machine learning, prognosis, model prediction

INTRODUCTION

Fluid resuscitation is the basic treatment for sepsis. However, in recent years, clinicians have found that an inappropriate infusion of large amounts of fluid may cause volume overload, which has become an independent risk factor for disability and death during critical illness (1–4). In the early treatment of sepsis, to improve organ perfusion, fluid therapy should be performed in a timely manner, but continuous positive fluid balance is not advocated. With the infusion of a large amount of fluid, a positive balance may cause treatment-related damage, such as capillary leak syndrome and pulmonary edema (5). In fact, an increasing number of studies suggest that having a negative fluid balance during the late treatment of sepsis can improve overall prognosis (6). However, how to prevent and address the fluid balance problem caused by volume therapy in clinical practice has become a problem that puzzles clinicians. Recently, physicians have used several methods to investigate resuscitation strategies for septic shock patients. Ma et al. (7) used finite mixture modeling and K-means clustering to identify subclasses of septic shock and dynamic treatment regime was used to give customized fluid volume and norepinephrine dose prescription for each patient. Lu et al. (8) explore the sensitivity of Dueling Double Deep Q-Networks to data preparation and modeling decisions in the context of hemodynamic management in septic patients. In this paper, we would like to explore fluid therapy balance strategies for ICU sepsis patients based on SARSA Algorithm, Q-learning and DQN model.

OBJECTS AND METHODS

Patient Enrollment, Data Extraction, and Interventions

Inclusion and Exclusion Criteria

Patients who were diagnosed with sepsis between January 2016 and April 2020 at the Department of Intensive Care, Peking Union Medical College Hospital, were included in this study. Informed consent was given by patients (or their legally authorized representative/next of kin if the patients were dead) before any data extraction. Patients were excluded from this study if they met any of the following criteria: (1) they were younger than 18 years old or (2) they were admitted to the ICU for fewer than 24 h.

Definition of Sepsis and Initial Time of Fluid Resuscitation

The initial timing of fluid resuscitation is when patients are diagnosed as "sepsis". Specifically, definition of sepsis is the time when both "infection is diagnosed" and "new onset of ΔSOFA $\geq$ 2" are reached (9). Infection was defined as follows: (1) if an antibiotic was used first, the etiology was obtained within 24 h after antibiotic treatment OR (2) if the etiology was noted first, the antibiotic was used within 72 h.

Data Extractions

Data were extracted from the Peking Union Medical College Hospital Intensive Care Medical Information System and Database (PICMISD). The parameters extracted from the database included demographics, vital sign data collected by bedside monitoring, laboratory examination data, blood gas analysis, microbiological examination results, antibiotic usage and total fluid balance. Notably, data de-identification was performed before any further analysis.

Ethical Approval

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. All the experimental protocols were approved by the Institutional Research and Ethics Committee of Peking Union Medical College Hospital, which approved this study for human subjects (No. SK-1241).

Data Collection and Cleaning Strategy

A total of 2,705 patients were included based on sepsis 3.0. The average ICU time was 4.3-day, as calculated by PICMISD. We included patient data over each 6 h period up to 108 h (4.5 days), for 18 periods in total. For patients who transferred out of the ICU or died within 108 h, the last period was their last record in the ICU. The values of each feature are the average value for each period. The final outcome was patient death or survival over the last period, that is ICU survival or death. Death includes clinical death and the withdrawal of treatment without any further action. Each 6 h is identified as a block, which is

abbreviated as "Bloc". Because the laboratory features do not need to be measured frequently, the missing values were forward-filled. Missing values of other features with <30% missing rate were filled by KNN. Twenty-seven features were used for modeling, including 25 state features (bloc, vital sign, laboratory examination, blood gas and demographics), one action feature (fluid balance), and one outcome feature. We included 2,443 sepsis patients, with 2,095 survivors and 348 non-survivors. Each group of data includes all 27 features for one period for one patient. Each patient could contribute 18 groups of data at most depending on their ICU time. There were 31,425 groups of data generated from 2,443 patients (the data selection strategy is shown in **Figure 1**). The features and outlier manipulation criteria are shown in **Table 1.**

State, Action, and Rewards

Q-Learning is a basic form of Reinforcement Learning which uses Q-values (also called action values) to iteratively improve the behavior of the learning agent. Q-Learning technique is an off-Policy technique and uses the greedy approach to learn the Q-value. SARSA technique, on the other hand, is an On Policy and uses the action performed by the current policy to learn the Q-value. DQN model is based on Q-learning, which we use to uncover the relationship between states and actions, obtain new knowledge from existing data and find the best treatment plan to predict the reasonable range of fluid balance.

The purpose of reinforcement learning is to obtain an optimal policy for a specific problem to maximize the reward obtained under a given strategy. A strategy is mapped from state to action, which can determine what behavior to choose under a certain state and then enter the next state. The transition in the state is equivalent to adding a reward to the Markov decision process. Hence, the next state is related not only to the current state but also to the current action. Every decision consists of a status, behavior and reward. Given that the prediction is the interval of fluid balance, which can be divided into five intervals, we used the value iteration method of reinforcement learning.

In this research, each data point is a group of states, which includes 25 features except the balance and outcome. The action refers to how much liquid should be given under a certain state, which is divided into five groups by 0–20, 20–40, 40–60, 60–80, and 80–100% of the percentiles of fluid balance. Each group of actions and the corresponding balance is shown in **Table 2**. "Rewards" refers to a survival-derived score. At the terminal timestamp of each patient, we issued a reward of +15 if they survived their ICU stay and a reward of−15 if they died. The rewards were all 0 at the other timestamps.

Algorithm Description

All the patients were divided randomly 8:2 into a training set and test set, corresponding to 2,095 (25,149 blocs) in the training data and 348 patients (6,326 blocs) in the test data. We needed to evaluate the expected reward when performing an action in a certain state; the higher the expected reward is, the better the effect of performing this action. The reward is called $Q(s, a)$, with s being the current state and a being the performing action. $Q(s, a)$ is also the q-value performing action a in state s. All $Q(s, a)$ are initialized as 0, and $Q'(s, a)$ is updated by performing different actions under different states and obtaining rewards from different next states. Then, the model can learn the possibilities of different $Q(s, a)$ under a certain state and choose the action with the highest possibility as the next action.

There are two methods to update $Q'(s, a)$: SARSA and Q-learning (10).

The formula for updating the $Q'(s, a)$ in SARSA is as follows:

$$\begin{aligned} Q'(s,a) &\leftarrow Q(s,a) + \alpha[r + \gamma Q(s',a') - Q(s,a)] \\ s &\leftarrow s'; a \leftarrow a' \end{aligned}$$

The formula for updating the $Q'(s, a)$ of Q-learning is as follows:

$$\begin{aligned} Q(s,a) &\leftarrow Q(s,a) + \alpha[r + \gamma max_{a'} Q(s',a') - Q(s,a)] \\ s &\leftarrow s' \end{aligned}$$

where r is the reward of the current action and α and γ are the parameters of 0–1.

When updating the $Q'(s, a)$, SARSA chooses a' from all possible next actions to the next state s' and obtains $Q'(s', a')$. Then, it performs a' in state s to state s'. $Q'(s, a)$ will be updated in the next iteration from one next action in state s'. Q-learning updates $Q(s, a)$ with the highest $Q'(s', a')$ in state s, but the next action will be chosen in another way in the next iteration rather than using the a' of $Q'(s', a')$. In general, the a' in state s is chosen randomly; in this case, we could find out how much balance (a_1) the patients have in state s_1, then enter state s_2, but if the action is random (such as a_2), and all patients in state s_1 did not have the balance a_2 in the existing data, we would not know what the next state could be. Thus, the next state should be chosen from the given sequence of data. The updated strategy for $Q(s, a)$ with SARSA is exactly the same as the actual implemented

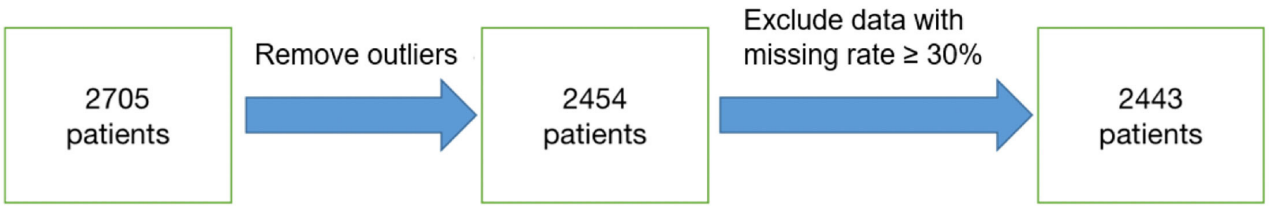


FIGURE 1 | Data selection strategy.

TABLE 1 | Missing rate and outlier manipulation criteria of the modeling features.

Feature	Description	Missing rate	Outlier manipulation method
Invasive mean pressure (mmHg)	Vital sign	0.033	Exclude data with values below 0 and the value is 0 but not dead
Invasive systolic blood pressure (mmHg)	Vital sign	0.034	Exclude data with values below 0 and the value is 0 but not dead
Invasive diastolic blood pressure (mmHg)	Vital sign	0.034	Exclude data with values below 0 and the value is 0 but not dead
Temperature (°C)	Vital sign	0.004	-
Breathe rate (bpm)	Vital sign	0.0002	Exclude data with values above 100 and the value is 0 but not dead
Oxygen concentration (%)	Vital sign	0.204	Set the values below 21 to 21. Exclude data with values above 100
Perfusion index	Vital sign	0.004	Exclude data with values above 50
CVP (mmHg)	Vital sign	0.250	Exclude data with values below or equal to 0
SPO_2 (%)	Vital sign	0.001	Exclude data with values of 0 or greater than 100
Heart rate (bpm)	Vital sign	0.0002	Exclude data with values of 0 but not dead
White blood cell ($\times 10^9$/L)	Laboratory examination	0.581	-
Neutrophilic granulocyte percentage (%)	Laboratory examination	0.583	Exclude data with values of 0
Hemoglobin (g/L)	Laboratory examination	0.581	-
Blood platelets ($\times 10^9$/L)	Laboratory examination	0.581	-
Creatinine (mmol/L)	Laboratory examination	0.675	-
Total bilirubin (mmol/L)	Laboratory examination	0.725	-
pO_2 (mmHg)	Blood gas	0.074	-
pCO_2 (mmHg)	Blood gas	0.074	-
BE	Blood gas	0.096	-
pH	Blood gas	0.074	Exclude data with values below 6.7
Lactate (mmol/L)	Blood gas	0.074	Exclude data with values above 30
Gender	Demographics	-	-
Age (yrs)	Demographics	-	-
Weight (kg)	Demographics	-	-
bloc		-	-
Balance (mL)	Output volume-input volume	-	Exclude data with outputs or inputs below 0 or above 5,000 and empty values
Outcome	Dead or survived	-	-

TABLE 2 | Fluid balance as the division of actions.

Actions	Action 0	Action 1	Action 2	Action 3	Action 4
Fluid balance intervals (mL)	<-110.68	−110.68 to −45.68	−45.68 to −0.67	−0.67 to 45.00	>45.00

strategy, which is called on-policy, while inconsistent with Q-learning, which is called off-policy. SARSA is aimed at learning the characteristics of the original data, and Q-learning tends to discover new strategies (11).

Data Evaluation

First, we used SARSA to learn the characteristics of the original data to then obtain the relationship between reward and mortality to evaluate whether the rewards were reasonable. $Q(s, a)$ is the expected reward. The relationship between the expected reward and mortality could be acquired by calculating the mortality rate of each data point in $Q(s, a)$. Ideally, the higher the expected reward is, the lower the mortality rate. We used the training set to build the SARSA model. Since the state cannot be exhaustive, we used a function (neural network) as the state to build a reinforcement learning model. The neural network consists of 1 input layer, two hidden layers, and one output layer. The input layer consists of 25 features (25 nodes). Each hidden layer has 128 hidden nodes. The output layer is the probability of $Q(s, a)$ corresponding to the five action categories, and softmax is used to select the highest probability action as the final output.

Each node of the neural network represents a function and includes weights w and biases b. The weights are different in different nodes, and the biases are the same in the same layer and different in different layers. w and b are initialized as randomly generated decimals of a normal distribution with a mean value of 0 and a standard deviation of 1, generally within the interval $[-1,1]$. Due to the size of the training set, data are usually trained in small batches in a fixed-size sample that is randomly selected from all training data for each iteration. The sample size is called the batch size, and the batch size of this project is 32. First, the network receives the input and calculates using the received input and parameters, comparing the estimated output with the real output, it obtains the mean square error, and then it updates the parameters of each node according to the error. The learning rate for updating the parameters is 0.0001 to ensure

that the amplitude of each update parameter will not be too large so that the current parameters can gradually approach the best parameters. Normally, the error will keep decreasing during training, and the decrease will be fast and then slow. When the error becomes stable as the training times increase, the training can be stopped to avoid overfitting.

Figure 2 shows the approach of SARSA updates $Q(s, a)$. In the project, there are five actions, a_1, a_2, a_3, a_4, a_5. Suppose the states are s_1,s_2,s_3.... Each data point includes a state and action group. The updating order is the same as the data order. As shown in **Figure 2**, suppose the first state is s_1, the action to be performed is a_2, and we update $Q(s_1, a_2)$. With first input s_1, we could have five Q values of five actions, then we could obtain s_2 as the next state and a_1 as the next action, using $Q(s_2, a_1)$ to update $Q(s_1, a_2)$. Then, we would update $Q(s_2, a_1)$... in the same way until the end of the iterations.

Balance Prediction

Next, we used deep Q-learning (DQN) based on Q-learning to uncover the relationship between states and actions, obtain new knowledge from existing data and find the best treatment plan to predict the reasonable range of fluid balance. We used the training set to train the model and the test set to make predictions on the trained model.

Figure 3 describes the approach of DQN updates with $Q(s, a)$. The updating order of $Q(s, a)$ is also the same as the data order; unlike SARSA, DQN uses the highest $Q(s, a)$ among five $Q(s, a)$ to update $Q(s_1, a_2)$. Suppose the highest $Q(s, a)$ is $Q(s_2, a_4)$, and we use $Q(s_2, a_4)$ to update $Q(s_1, a_2)$.

The principle of updating $Q(s, a)$ with DQN and Q-learning is the same. Both use the max method. However, there will be an overestimation problem. If the data are noisy, the best $Q(s, a)$ is not the largest $Q(s, a)$. There may be biases after each iteration, and the biases will approach Q with greater values (12).

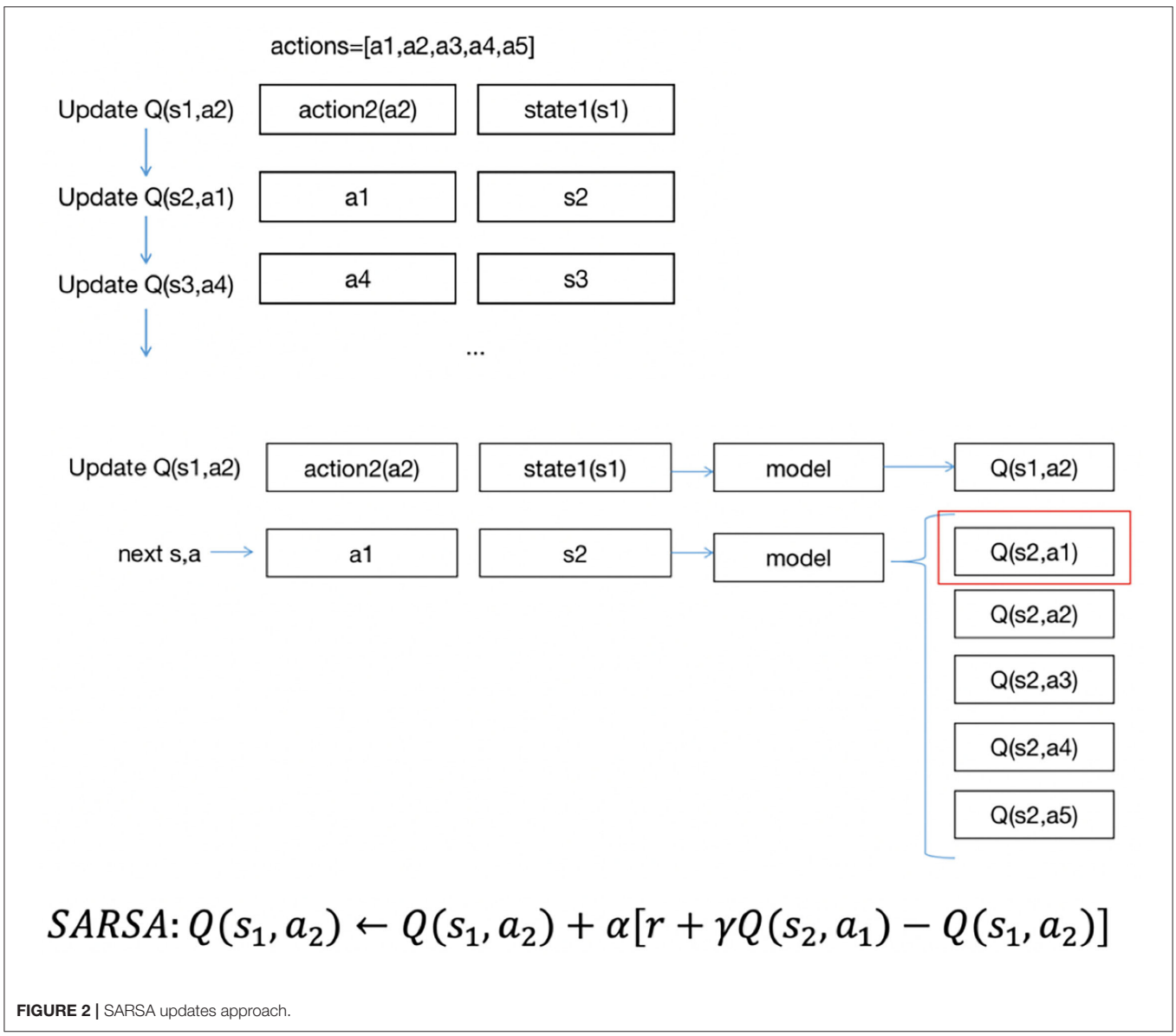


FIGURE 2 | SARSA updates approach.

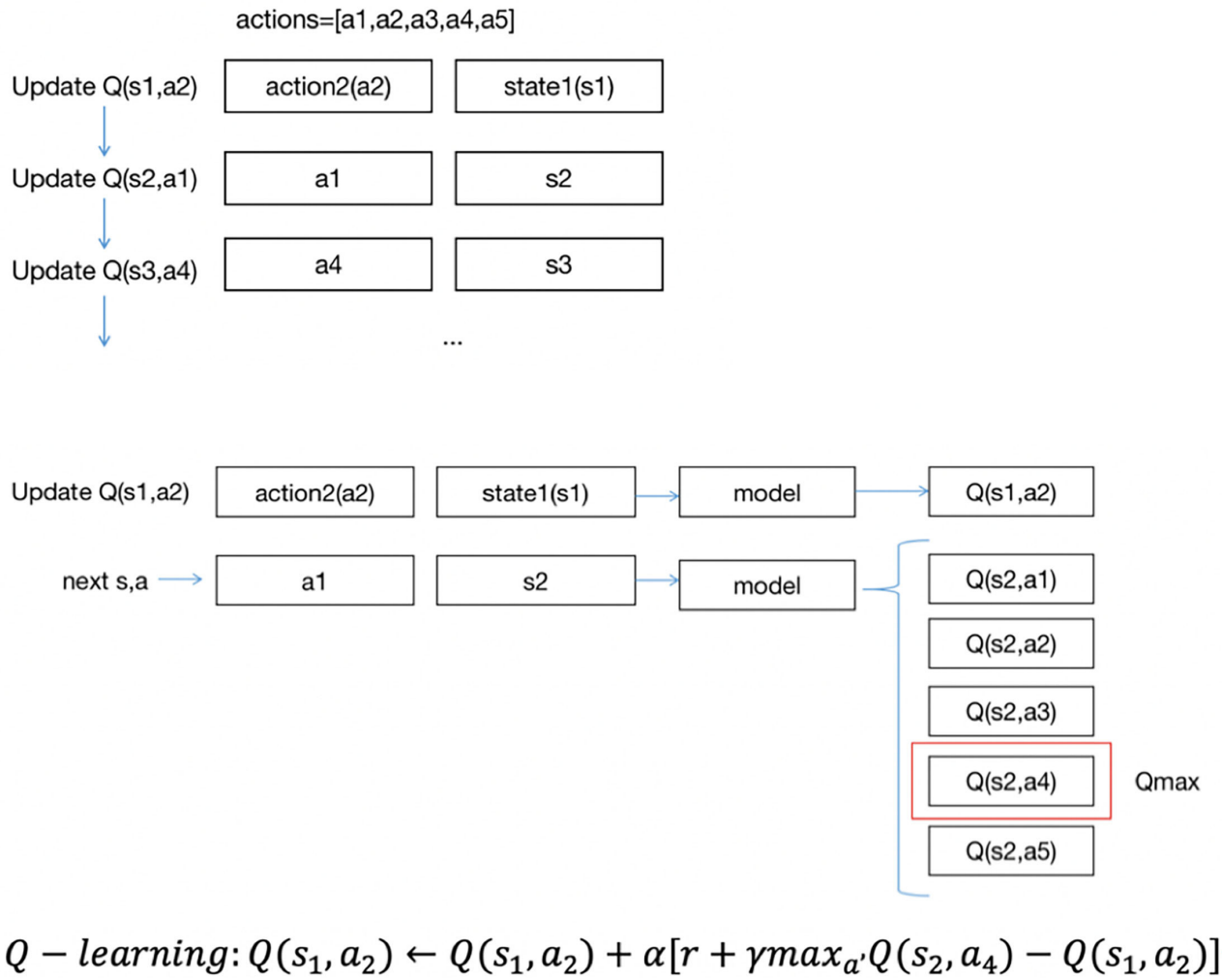


FIGURE 3 | DQN updates approach.

To solve the problem, we use double deep Q-learning (DDQN). The method of updating $Q(s, a)$ with DDQN is as follows:

$$\begin{aligned} Q(s_t, a_t) &\leftarrow Q(s_t, a_t) + \alpha[R_{t+1} + \gamma Q'(s_{t+1}, a) - Q(s_t, a_t)] \\ a &= max_a Q(s_{t+1}, a) \end{aligned}$$

The method of calculating losses with DQN and DDQN is shown in **Figure 4**.

θ is the parameter of the neural network; the method of obtaining Q is called selection, and the calculation loss is called an evaluation. DDQN implements the two methods with two Q networks, avoiding the risk of using the same Q network.

We define two identical structured Q networks, the primary QN and target QN, train with the primary QN and evaluate with the target QN (13). The calculation process is as follows:

(1) Use the primary QN to select the action with the highest $Q(s, a)$ in s_{t+1}, where $Q(s, a)$ is called q_1;
(2) Use target QN to evaluate the $Q(s, a)$ of the action, and the evaluated $Q(s, a)$ is called q_2;
(3) Set the difference between q_2 and q_1 as the loss, update the parameters of the primary QN with back propagation, and do not update the parameters of target QN;
(4) Repeat (1) (2) (3) until convergence.

Algorithm Evaluation

The algorithm evaluation method compares the original average reward and the average reward calculated by the model on the test set. If the average reward calculated by the model is higher than that of the original test set, we can argue that the model has an effect on optimizing treatments. However, because the model is built on the training set, we cannot ensure that the distribution (probability of occurrence) of the test set is the same as that of the training set, and it is inaccurate to use the mean prediction rewards for comparison. To ensure that the two groups of data

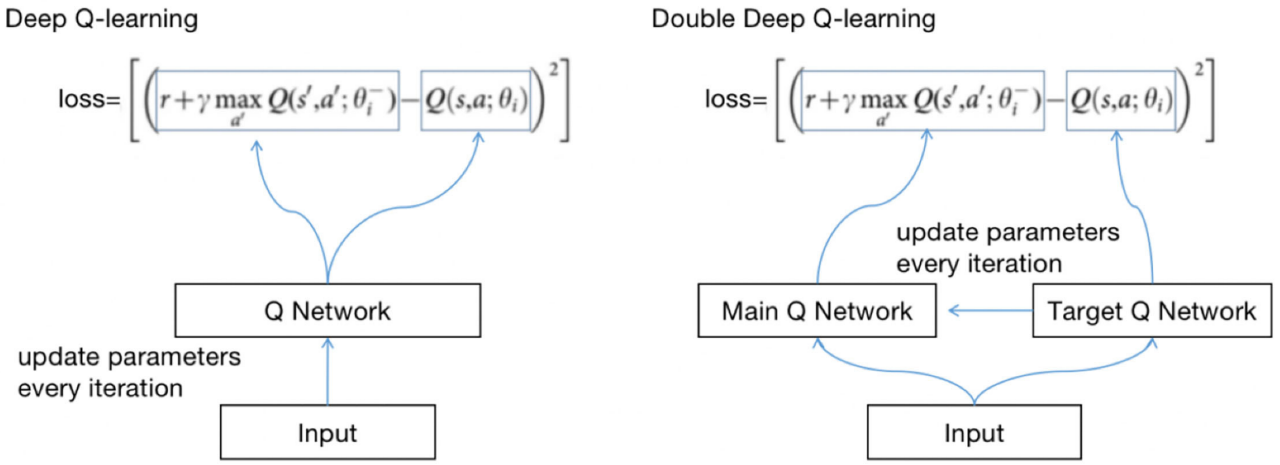


FIGURE 4 | Calculating loss with DQN and DDQN.

are unbiased, the inverse probability score (IPS) is generally used for approximating the test set to the training set distribution.

The formula of the IPS is as follows:

$$V_{IPS}^{\pi} = \frac{1}{|N|}\sum_{i=1}^{N}\frac{A_i r_i}{\pi_i(x_i)} \quad (1)$$

where A is whether an event occurs or not, r is the reward, and $\pi_i(x_i)$ is the probability of an event occurring. If the probability on the test set is lower than that of the training set, as the denominator, r will be multiplied by a multiple greater than 1, which is the inverse. If the probability of an event occurring is correct, the test set will be unbiased. However, there are two problems. First, we cannot ensure that the probability is completely correct. Second, the variance will be too large if the probability is very small.

We can also use regression-based model $m(x_i)$ for prediction, and the formula is as follows:

$$m(x_i) = E(r|A=1,x) \quad (2)$$

where A indicates whether AI policy is consistent with clinical policy, 1 indicates consistency and 0 indicates inconsistency.

$$V_{model} = \frac{1}{N}\sum_{i=1}^{N}[A_i r_i + (1-A_i)m(x_i)] \quad (3)$$

If $A=1$, $(1-A_i)$ is 0, and according to the formula, the original reward will be used. If $A=0$ and $A_i r_i$ is 0, the predicted reward will be used. Similarly, if the model is correct, the result will be unbiased.

A double-robust estimator is the combination of IPS and a regression model. The formula is as follows:

$$V_{DR}^{\pi} = \frac{1}{|N|}\sum_{i=1}^{N}[\frac{A_i(r_i - m(x_i))}{\pi_i(x_i)} + m(x_i)] \quad (4)$$

OR:

$$V_{DR}^{\pi} = \frac{1}{|N|}\sum_{i=1}^{N}[\frac{A_i r_i}{\pi_i(x_i)} - \frac{A_i - \pi_i(x_i)}{\pi_i(x_i)}m(x_i)] \quad (5)$$

where $m(x_i)$ is the model, which is used for predicting the reward. If the model is correct, according to formula 4, the model is able to predict r_i, and if $r_i - m(x_i) = 0$, $V_{DR}^{\pi} = \frac{1}{|N|}m(x_i)$, the result is unbiased. If the model is not correct, the probability of an event occurring is correct. According to formula 5, $V_{DR}^{\pi} = \frac{1}{|N|}\sum_{i=1}^{N}\frac{A_i r_i}{\pi_i(x_i)}$, and the result is also unbiased. In other words, at least one of the models and the event probability estimation is correct, and the result is unbiased (doubly robust) (14–17).

Model Validation

Patient data from July 2019 to April 2020 were extracted from PICMISD as the validation set. We compared the predictions and actual clinical conditions and prognoses of four different types of negative fluid balance and fluid infusion in the validation set. The preprocessing method of the validation set is the same as that used for the data collection and cleaning strategy in Section 2. Lastly, the validation set included 399 sepsis patients (representing 5269 time-blocks) were used for verification, with 357 survivors (including 4677 time-blocks), and 42 non-survivors (including 592 time-blocks). We divided the predicted and clinical 5-group fluid balance into two groups based on the total amount of fluid (positive or negative balance). The previous categories 0, 1, and 2 correspond to the new category 0, indicating negative fluid balance (negative balance), and categories 3 and 4 correspond to the new category 1, indicating fluid infusion (positive balance). In addition, the actual negative fluid balance and fluid infusion were determined according to the actual clinical PICMISD records. We calculated the morality of the four combinations of prediction and clinical results. The mortality rate is the lowest when the predicted fluid management strategy is the same as the actual strategy.

Statistical Analysis

All statistical analyses were performed using SPSS Statistics for Windows (version 19.0, IBM Corporation, Armonk, NY) and R 3.4.3. Student's *t*-test or the Wilcoxon rank sum test was

used to compare continuous variables. The categorical data were compared using the chi-square test or Fisher's exact test. The Kruskal-Wallis test was used to compare differences in continuous parametric variables with abnormal distributions. Differences in the variables between the groups were considered statistically significant at the a p level of < 0.01.

RESULTS

General Description

Table 3 shows the significant difference between each feature of the training set and test set. The p-values of all the features are greater than 0.01, which means that there is no significant difference between the two datasets.

Data Distribution

As shown in **Figure 5**, from left to right, the first row shows the fluid balance distribution, action distribution and reward distribution of the training set based on the frequency. The test set shown in the second row follows the same distribution. Each data point represents the data from each period for each patient. The balance distribution figure shows that most patients had a balance of ±500, with the highest being 1608.92 and lowest being −1097.58, which was similar to those in the test set. The action distribution figure shows that although the balance of each patient for each period is mostly different, the percentage of each action in the training set and test set are basically the same. The number of each group of actions in the training set and test set are the same. In the reward distribution figure on the training set, the rewards of 2,095 individuals are alive, and 348 individuals died. In addition, most of the rewards were 0. The proportion of rewards in each part of the test set is basically the same as that of the training set. The results above show that the distributions of the training set, test set, and validation set are basically the same.

Prediction of SARSA

Each training dataset includes a group of states and actions $Q(s, a)$. **Figure 6** describes the mortality and expected reward of the training set. Clearly, the higher the expected reward is, the lower the mortality rate, and the reward value is properly set.

Prediction of Q-Learning

Figure 7B shows the predicted action distribution of the test set. Action 2 has the highest amount among the five actions, and the corresponding balance is −45.68 to −0.67. Action 0 and action 4 have the lowest amount, which corresponds to too little and too

TABLE 3 | Comparison of 27 features of the training set and test set.

Features*	Training set median (25th, 75th)	Test set median (25th, 75th)	Validation set median (25th, 75th)	p-value (training vs. test set)	p-value (training vs. validation set)
Invasive mean pressure	88.33 (81.56–95.57)	89.25 (81.76–97.12)	83.74 (78.0–89.9)	0.1320	0.1320
Invasive systolic blood pressure	131.57 (120.47–143.62)	132.91 (121.83–145.56)	122.56 (112.02–135.0)	0.2992	0.2992
Invasive diastolic blood pressure	67.33 (60.75–74.22)	68.45 (61.41–76)	64.29 (58.79–69.58)	0.0144	0.0144
Temperature	37 (36.5–37.5)	37 (36.55–37.5)	37.0 (36.5–37.5)	0.2956	0.2956
Breathe	18.08 (15.86–20.83)	18.12 (15.71–21.1)	17.33 (15.38–20.2)	0.4522	0.4522
Oxygen concentration	31 (28–38.75)	31 (27.76–39.21)	36.29 (30.53–43.12)	0.1943	0.1943
Perfusion index	1.5 (0.79–2.4)	1.6 (0.81–2.63)	1.13 (0.64–1.87)	0.0224	0.0224
CVP	8 (6.5–9.64)	8 (6.33–9.61)	8.33 (7.0–10.0)	0.1711	0.1711
SPO_2 (%)	98.64 (97.45–99.6)	98.38 (97.14–99.29)	98.4 (96.84–99.5)	0.0500	0.0500
Heart rate	92.86 (82.5–103.67)	93.45 (82.33–105)	94.56 (85.1–102.97)	0.1275	0.1275
White blood cell	11.8 (8.49–16.59)	11.22 (7.94–15.47)	11.61 (7.71–16.29)	0.2308	0.2308
Neutrophilic granulocyte percentage	86.1 (80.6–90.2)	86.2 (80.46–90.4)	88.05 (82.3–92.03)	0.0163	0.0163
Hemoglobin	96 (86–109)	97 (88–110)	91.0 (83.67–100.46)	0.0190	0.0190
Blood platelets	144 (89–208)	138 (80–208)	93.0 (63.67–140.92)	0.0660	0.0660
Creatinine	86 (60–138)	79 (55–139)	106.0 (79.0–164.0)	0.0221	0.0221
Total bilirubin	16.9 (11.4–30.9)	16.7 (11.2–30.1)	28.1 (15.3–59.49)	0.2970	0.2970
Lac	1.3 (0.9–1.83)	1.30 (1–2)	1.6 (1.13–2.59)	0.3278	0.3278
pO2	92.80 (79.3–111)	92.91 (79.5–110)	94.72 (79.26–117.0)	0.4295	0.4295
pCO2	39.05 (35.8–42.6)	39.3 (35.95–43.1)	39.85 (36.26–43.0)	0.3987	0.3987
BE	3.03 (0.4–5.47)	3.17 (0.6–5.9)	3.0 (0.0–5.6)	0.4494	0.4494
pH	7.45 (7.42–7.48)	7.45 (7.41–7.48)	7.44 (7.41–7.47)	0.3254	0.3254
Age	62 (48–70)	62 (50–70)	59.0 (50.0–68.0)	0.2671	0.2671
Weight	65 (58–75)	65 (58–75)	65.0 (60.0–74.0)	0.4209	0.4209
Bloc	8 (4–13)	8 (4–12)	9.0 (4.0–13.0)	0.1800	0.1800
Fluid balance	−20.83 (−90.66–32.87)	−19.24 (−90.25–37.31)	−31.19 (−109.05–36.81)	0.1606	0.1606

**All parameters do not obey normal distribution.*

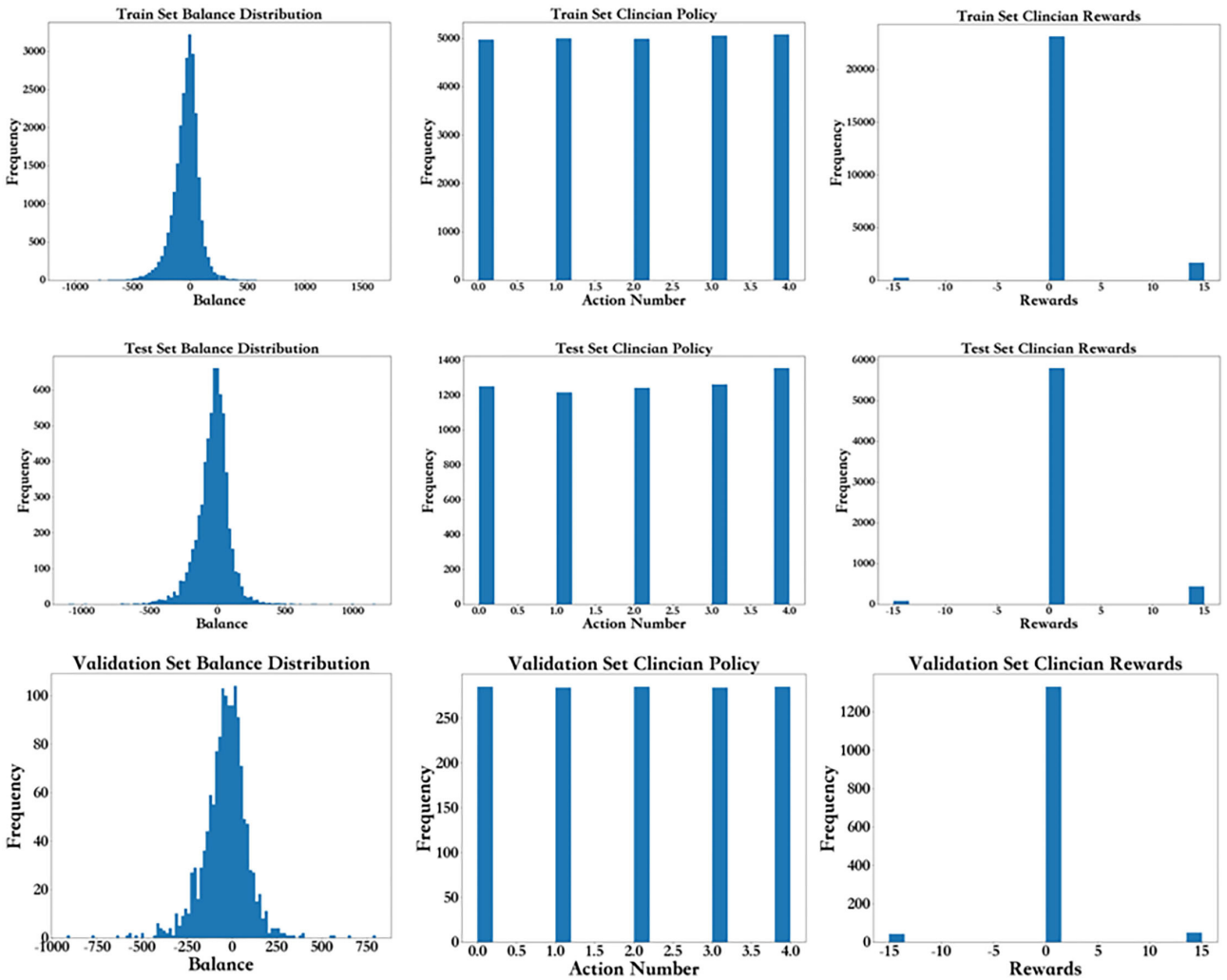


FIGURE 5 | Frequency distribution for the balance distribution, action distribution, and reward distribution of the training set, test, and validation set.

much balance. In comparing the origin distribution in **Figure 7A**, the amount of balance that was too high and too low decreased significantly. **Figure 7C** shows the APACHE II score of these different actions.

Figure 8 shows the relationship between fluid balance difference and mortality. We used the median fluid balance interval as the predicted value, and the difference was the predictive value minus the actual value. The closer the model is to the clinical fluid balance difference, the lower the mortality rate. This trend shows that the model is able to conclude clinical rules. However, a fluid balance difference that is too high or too low is not good for patients. Mortality is higher in patients with a lower fluid balance difference than in those with a higher fluid balance difference. The supplemental figure shows the predicted - real fluid balance difference and APACHE II score.

Evaluation

A double-robust estimator was used to calculate the average expected reward $[Q(s, a)]$ of the original test set, and the average predicted reward was calculated by the Q-learning model. The result is shown in **Table 4.** The results show that the average reward keeps increasing as the iterations increase. The increase is significant at the beginning, while the number of iterations rises from 20,000 to 30,000. Compared to the first 10,000 iterations, the increase in the average reward is much lower and could gradually become stable in future iterations. There could also be a risk of overfitting. Thus, we chose the model with 30,000 iterations as the final model.

Validation

As shown in **Figure 9A**, patient mortality was lowest when negative fluid balance was predicted to be the same as clinical in both the validation set and the test set. The mortality is lower when the reinforcement strategy is the same as the clinical strategy. When the reinforcement strategy is different from the clinical strategy, it is more serious to predict negative fluid balance as fluid infusion than to predict fluid infusion as negative fluid balance, and the mortality of the former is lower than that

of the latter. The results are shown in **Figure 9B**, which show the high value of the model in reality.

DISCUSSION

Our study indicates that the use of reinforcement learning methods can clearly predict a patient's future liquid treatment strategy. To our knowledge, clinical decision-making is a process that usually not gives immediate feedback. A certain treatment strategy may improve or worsen the state of illness. However, it does not mean the treatment prior to change of the illness is right or wrong. In clinical practice, a series of treatment procedures work together to accomplish the therapeutic effect. Hence, clinical decision-making in ICU is not a simple static classification problem. Instead, it is a dynamic process which we should learn through environmental rewards and punishments.

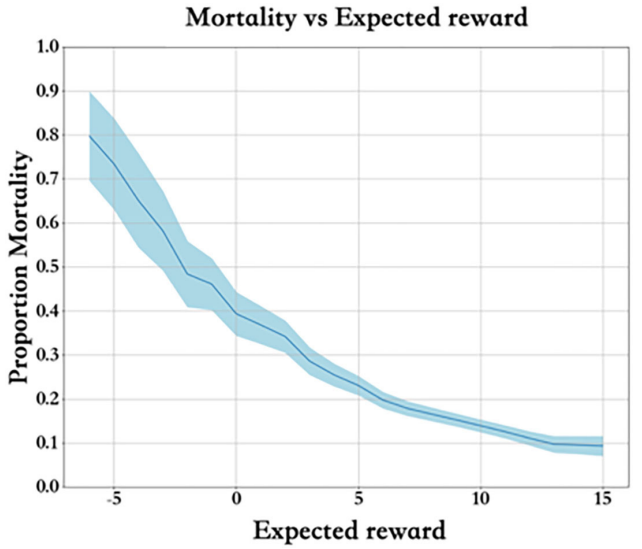


FIGURE 6 | Mortality and expected reward.

Under this circumstance, reinforcement learning should be used to maximize rewards or achieve specific goals through learning strategies in the interactive process of the environment. The SARSA model can be used to simulate the equation between expected mortality and actual mortality during the liquid treatment process. The Q-learning model shows that as the model prediction and actual intake and output become closer, the mortality rate decreases. Additionally, if the intake and output are too high or too low which caused abnormal fluid imbalance, the mortality rate would be higher. Patients with higher positive fluid imbalance may have higher mortality than those with a higher negative fluid imbalance. We used a double-robust estimator to calculate the average expected reward of the test set in the Q-learning model after training 30,000 times as the final prediction model. Using a validation data set, the results suggest that if the model predicts that the patient should be dehydrated while the patient is dehydrated under actual treatment, the fatality rate is significantly lower compared with other circumstances. In particular, these data are from the first 108 h of entering the ICU ward, which suggests that the resuscitation is not a positive mass fluid resuscitation. This is worth thinking about for the clinicians.

Fluid overload is related to the prognosis of severely ill patients with septic shock and/or acute respiratory distress syndrome (ARDS). Excessive fluid load may lead to a vicious cycle in which interstitial edema causes organ dysfunction, leading to fluid accumulation, organ edema and dysfunction. All the potentially harmful consequences of fluid overload in different end-organ systems have an impact on patient morbidity and mortality. Therefore, although infusion is advocated in early resuscitation strategies, the side effects of inappropriate or excessive infusion are increasingly recognized by practitioners. Fluid therapy can be considered a double-edged sword. In 2000, a retrospective cohort study conducted by Alsousand et al. indicated that patients with a negative fluid balance for at least one day during the first 72 h of septic shock had a better prognosis (OR 5.0; 95% CI 2.30–10.9, $p < 0.001$) (18). In 2006, the FACTT study showed that although negative fluid balance had no effect on the mortality rate, it can significantly reduce the time of mechanical ventilation and the ICU stay in critically ill patients (19). Since 2006, more studies have focused on

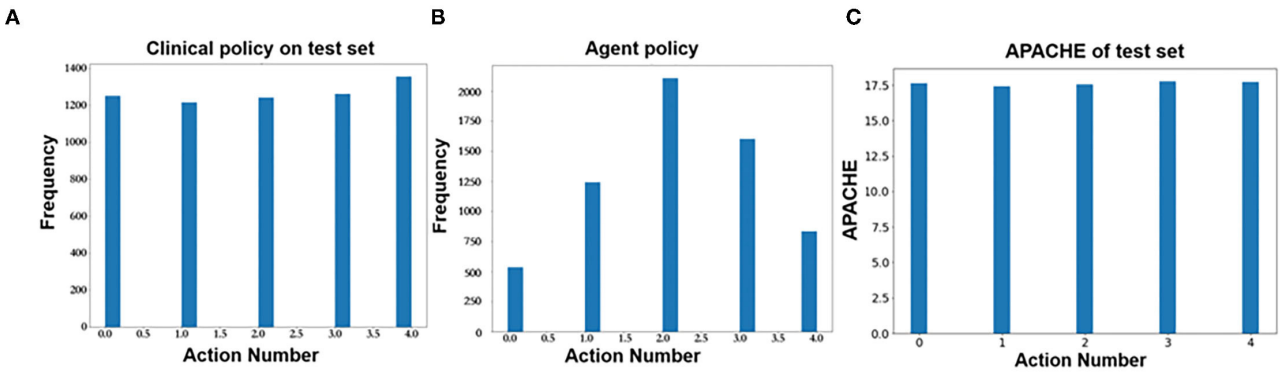


FIGURE 7 | (A) Action distribution of clinical policy on the test set. **(B)** Action distribution of AI policy on the test set. **(C)** Action distribution of APACHE II on the test set.

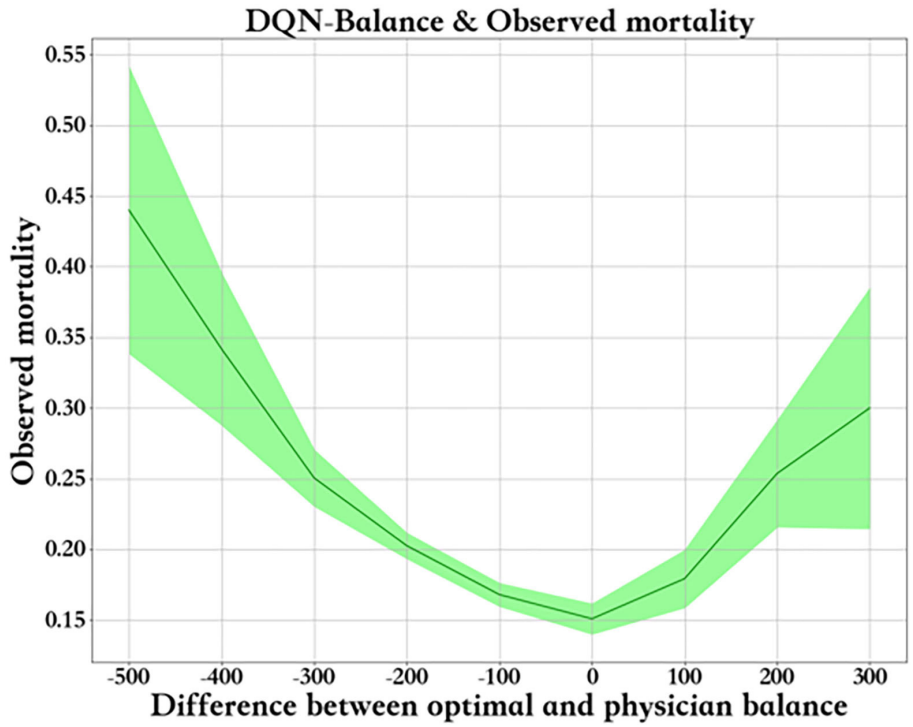


FIGURE 8 | Predicted - real fluid balance difference and morality.

the relationship between fluid balance and mortality in sepsis, suggesting the disadvantages of positive fluid balance for patients with sepsis from the perspective of evidence-based medicine (1–4). A recent systematic review involving fluid balance and prognosis in critically ill children suggests that after initial resuscitation, these children may develop edema and progress to fluid overload. More evidence shows that fluid overload causes more complicated treatment and serious outcomes, which could increase morbidity and mortality (20). Therefore, when treating patients with sepsis, we should be alert to the "reinjury" caused by excessive volume load, and more methods should be used to identify and assist in decision-making about fluid therapy strategies. In particular, our research has found that early infusion does not require a large amount of fluid infusion for the ICU admission in the real world. If fluid therapy can be performed more appropriately and a negative balance can be reached earlier, patients will ultimately benefit. For critically ill patients, at the beginning of ICU admission, it is necessary to choose an assessment every 6 h. Regarding the liquid usage within these 6 h period, actual fluid balance rather than the actual amount of input and output is the core of the resuscitation treatment. Our study provided effective treatment decision-making recommendations that have good predictive performance based on fluid balance and each time interval. For example, patients with "recommended negative fluid balance but actual negative fluid balance" have a better prognosis, while those with "recommended negative fluid balance but actual positive fluid balance" have the worst prognosis.

TABLE 4 | Average expected reward.

Items	Average reward
Original data from test set	4.07
Q-learning model (3,000 iterations)	4.05
Q-learning model (10,000 iterations)	9.06
Q-learning model (20,000 iterations)	10.37
Q-learning model (30,000 iterations)	10.47

The average infusion volume of patients with severe infection and septic shock on the first day of admission to the ICU is lower than recommended by the Surviving Sepsis Campaign bundle. An infusion of more than 5 L of fluid on the first day of admission to the ICU is associated with a significant increase in the risk of death and a significant increase in hospitalization costs (21). Vincent and De Backer recently proposed a conceptual model for managing the shock state. This model is aimed at fluid management during the treatment of critical illness. The treatment of shock is divided into four phases: (1) recovery phase: the goal is to achieve the lowest blood pressure level sufficient to maintain life; (2) optimum phase: the goal is to increase cardiac output to meet the requirements of the body; (3) stable phase: the goal of this phase is organ support and avoiding complications; and (4) de-escalation phase: patients in this stage should gradually leave the ICU intervention measures. This approach formally puts forward the necessity of liquid de-escalation therapy (22, 23). On this basis, Monnet et al.

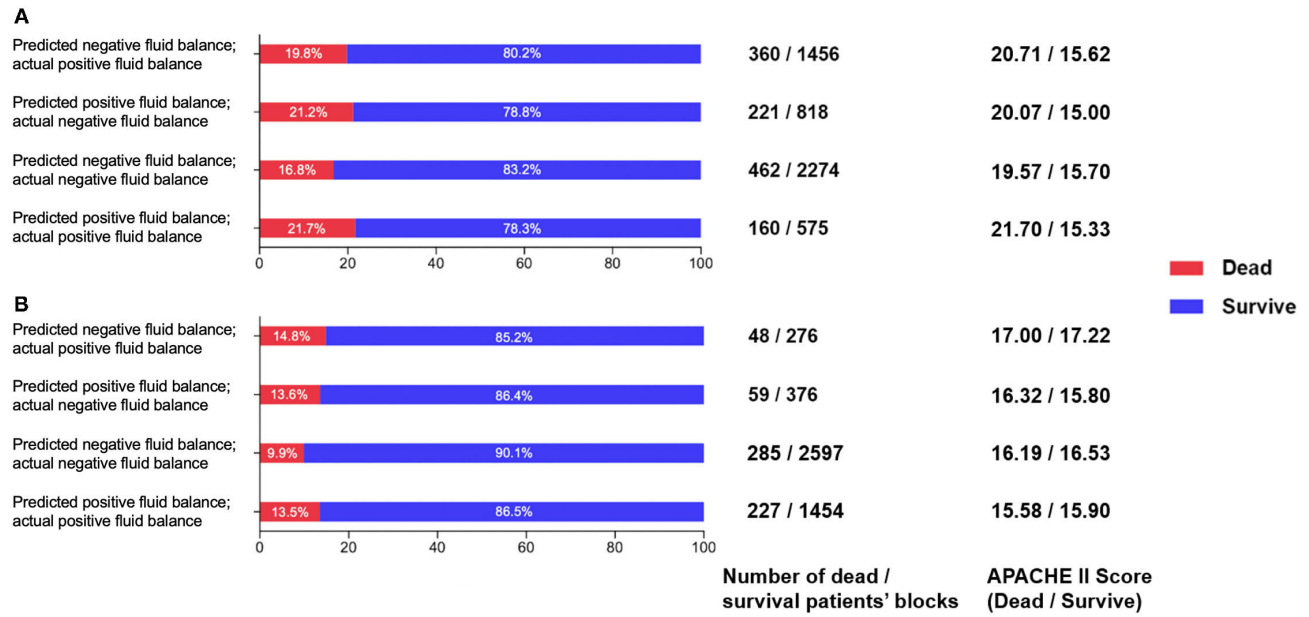


FIGURE 9 | (A) Comparison of survival and death patients' blocks in the prediction and clinical balance groups in the validation set. **(B)** Comparison of survival and death patients' blocks predictions and clinical balance in the test set.

(24) discussed different fluid management strategies, including early goal-oriented fluid management, late conservative fluid management, and late goal-oriented fluid removal management. In addition, the "Four-D" (medication, Dosage, Duration and Degradation) concept of fluid therapy is expanded. When treating patients with septic shock, four stages of fluid therapy should be considered to answer these questions. Doctors should be fully aware of the proper time to start or stop intravenous infusion, when to start reverse resuscitation or actively drain fluid, and when to stop reverse resuscitation. However, it is difficult to give a clear standard to explain how patients should implement fluid therapy accurately. These abstract concepts are not conducive to clinical implementation. Sometimes the success or failure of a treatment depends on the doctor's own experience and understanding of related theories. Recent studies have shown that achieving a negative volume balance in the ICU is associated with a reduction in 90-day mortality. An earlier negative fluid balance is associated with a reduction in mortality. Each liter of negative fluid balance increases the mortality rate (20). This finding shows that fluid treatment, especially the identification of the de-escalation stage, is of great significance. Our model proposes this possibility both theoretically and practically. Using the Q-learning model can provide us with the direction of liquid therapy for clinical reference. Our model really brings the theoretical idea of fluid resuscitation back to the scale of clinical operability, with the help of computer reinforcement learning.

However, we should never forget avoiding excessive negative fluid balance during treatment. Our goal for negative fluid balance is to remove excess volume in the interstitial spaces. However, the volume in the circulation must be removed first. When the interstitial fluid resorption rate (plasma refilling rate) is sufficient to prevent hypovolemia, hypotension does not occur. Currently, we do not completely know the lowest fluid resorption rate that may prevent hypotension. Studies have shown that in patients with severe negative fluid balance, increased fluid intake and urine output are related to a decrease in hospital mortality. However, achieving a more negative fluid balance compared to a mild fluid balance is not associated with reduced mortality (25). In addition to giving certain liquid treatment strategies, our model suggests that excessive negative fluid balance and positive fluid balance also bring side effects. Our model gives us a better basis for making choices, which gave a boundary between negative fluid balance and positive fluid balance. This is the contribution of machine learning to precision medicine.

Several limitations should be mentioned in this study. This study uses a single-center database. The sample size and the treatment stereotype of the treatment center may weaken the universality of the model. In particular, the weights of 18 blocks for each patient were equal in this study. At present, a method to solve the time series problem has been presented (26, 27). Further study, including temporal evolution along blocks of the fluid balance, can be performed. In addition, the amount and timing of negative fluid balance and positive fluid balance in this model cannot be completely calculated. Our finding can give directions for fluid therapy dependent on model predictions. Hence, the results given by the model should be combined with the clinical results. Any current medical applications of artificial intelligence cannot replace physician's medical decision making. Third, several confounding factors may influence the result of analysis. For example, usage of vasopressors and variation of cardiac functions may contribute to outcome and fluid balance. Patient's severity of illness is not invariable during

the treatment progresses. Moreover, the treatment of severe illness is very complicated. Although we considered the impact of disease severity when analyzing each bloc, it is impossible for us to update the value of the relevant scoring system every 6 h according to current data. The impact of any machine learning model on actual clinical conditions must be confirmed by BCT studies. Whether other interventions have also affected the process and conclusions of reinforcement learning, we may answer and solve them through more advanced methods if possible (28, 29).

CONCLUSION

This study proves that the reality of fluid therapy for patients with sepsis in the early stage of ICU admission is that a large amount of fluid infusion may not be required. If it can be converted to the negative fluid balance resuscitation phase as soon as possible, the patient's prognosis is better. Reinforcement learning methods were used to propose a possible predictive model for guiding the fluid therapy of patients with sepsis in ICUs.Our study presents a methodological model for fluid therapy. It is believed that machine learning will ultimately assist in clinical decision-making regarding the fluid therapy of critically ill patients in the future.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: This data was collected by Peking Union Medical College Hospital Intensive Care Medical Information System and Database (PICMISD), which is a protected database. If needed, we would like to disclose the data after the permission of Ethics Committee of Peking Union Medical College Hospital. Requests to access these datasets should be directed to LS, slx77@163.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Research and Ethics Committee of Peking Union Medical College Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YLo, WZ, and BD takes responsibility for the integrity of the work as a whole. LS is responsible for the study design and conception and drafted the manuscript. YLi, MS, and SZ were responsible for the study design and conception and for the data management and statistical analysis. LS, SL, XZ, and LW drafted the manuscript. All authors revised the manuscript for content. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by CAMS Innovation Fund for Medical Sciences (CIFMS) from Chinese Academy of Medical Sciences (No. 2021-I2M-1-062 & 056), National Key R&D Program from Ministry of Science and Technology of the People's Republic of China (No. 2021YFC2500800), Beijing Nova Program from Beijing Municipal Science and Technology Commission (No. Z201100006820126), China Health Information and Health Care Big Data Association Severe Infection Analgesia and Sedation Big Data Special Fund (No. Z-2019-1-001), China International Medical Exchange Foundation Special Fund for Young and Middle-Aged Medical Research (No. Z-2018-35-1902), Excellence Program of Key Clinical Specialty of Beijing for critical care medicine in 2020 (ZK128001), and Beijing Municipal Science and Technology Commission (No. Z201100005520051).

REFERENCES

1. Acheampong A, Vincent JL. A positive fluid balance is an independent prognostic factor in patients with sepsis. *Crit Care.* (2015) 19:251. doi: 10.1186/s13054-015-0970-1
2. Boyd JH, Forbes J. Nakada T-a, Walley KR, Russell JA: Fluid resuscitation in septic shock: a positive fluid balance and elevated central venous pressure are associated with increased mortality. *Crit Care Med.* (2011) 39:259–65. doi: 10.1097/CCM.0b013e3181feeb15
3. Smith SH, Perner A. Higher vs. lower fluid volume for septic shock: clinical characteristics and outcome in unselected patients in a prospective, multicenter cohort. *Crit Care.* (2012) 16:R76. doi: 10.1186/cc11333
4. Kelm DJ, Perrin JT, Cartin-Ceba R, Gajic O, Schenck L, Kennedy CC. Fluid overload in patients with severe sepsis and septic shock treated with early goal-directed therapy is associated with increased acute need for fluid-related medical interventions and hospital death. *Shock.* (2015) 43:68–73. doi: 10.1097/SHK.0000000000000268
5. Siddall E, Khatri M, Radhakrishnan J. Capillary leak syndrome: etiologies, pathophysiology, and management. *Kidney Int.* (2017) 92:37–46. doi: 10.1016/j.kint.2016.11.029
6. Dhondup T, Tien JCC, Marquez A, Kennedy CC, Gajic O, Kashani KB. Association of negative fluid balance during the de-escalation phase of sepsis management with mortality: a cohort study. *J Crit Care.* (2020) 55:16–21. doi: 10.1016/j.jcrc.2019.09.025
7. Ma P, Liu J, Shen F, Liao X, Xiu M, Zhao H, et al. Individualized resuscitation strategy for septic shock formalized by finite mixture modeling and dynamic treatment regimen. *Crit Care.* (2021) 25:243. doi: 10.1186/s13054-021-03682-7
8. Lu M, Shahn Z, Sow D, Doshi-Velez F, Lehman L-WH. Is Deep reinforcement learning ready for practical applications in healthcare? A sensitivity analysis of duel-DDQN for hemodynamic management in sepsis patients AMIA. *Annu Symp Proc.* (2020) 2020:773–82.
9. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA.* (2016) 315:801–10. doi: 10.1001/jama.2016.0287
10. Niranjan. *GARM: On-line Q-learning using connectionist systems*. Technical Report CUED/F-INFENG/TR 166. Cambridge (1994).
11. HV H. Double Q-learning. *Adv Neural Inf Process Syst.* (2010) 23:2613–21.
12. Hado van Hasselt AG, David Silver: Deep reinforcement learning with double Q-learning. *arXiv [Preprint]*. (2015) 1509:06461.
13. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nat Med.* (2018) 24:1716–20. doi: 10.1038/s41591-018-0213-5

14. Miroslav Dudik JL, Lihong Li. Doubly robust policy evaluation and learning. *arXiv [Preprint]*. (2011) 1103:4601.
15. Nan Jiang LL: Doubly robust off-policy value evaluation for reinforcement learning. *arXiv [Preprint]*. (2016) 1511:03722.
16. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res.* (2011) 46:399–424. doi: 10.1080/00273171.2011.568786
17. SNaPS JPH: Importance sampling policy evaluation with an estimated behavior policy. *arXiv preprint*. (2019) 1806:01347. doi: 10.48550/arXiv.1806.01347
18. Alsous F, Khamiees M, DeGirolamo A, Amoateng-Adjepong Y, Manthous CA. Negative fluid balance predicts survival in patients with septic shock: a retrospective pilot study. *Chest.* (2000) 117:1749–54. doi: 10.1378/chest.117.6.1749
19. Wiedemann HP, Wheeler AP, Bernard GR, Thompson BT, Hayden D. deBoisblanc B, et al. Comparison of two fluid-management strategies in acute lung injury. *N Engl J Med.* (2006) 354:2564–75. doi: 10.1056/NEJMoa062200
20. Alobaidi R, Morgan C, Basu RK, Stenson E, Featherstone R, Majumdar SR, et al. Association between fluid balance and outcomes in critically ill children: a systematic review and meta-analysis. *JAMA Pediatr.* (2018) 172:257–68. doi: 10.1001/jamapediatrics.2017.4540
21. Marik PE, Linde-Zwirble WT, Bittner EA, Sahatjian J, Hansell D. Fluid administration in severe sepsis and septic shock, patterns, and outcomes: an analysis of a large national database. *Intensive Care Med.* (2017) 43:625–32. doi: 10.1007/s00134-016-4675-y
22. Vincent J-L, De Backer D. Circulatory shock. *N Engl J Med.* (2013) 369:1726–34. doi: 10.1056/NEJMra1208943
23. Hoste EA, Maitland K, Brudney CS, Mehta R, Vincent JL, Yates D, et al. Four phases of intravenous fluid therapy: a conceptual model. *Br J Anaesth.* (2014) 113:740–7. doi: 10.1093/bja/aeu300
24. Malbrain MLNG, Van Regenmortel N, Saugel B, De Tavernier B, Van Gaal P-J, Joannes-Boyau O, et al. Principles of fluid management and stewardship in septic shock: it is time to consider the four D's and the four phases of fluid therapy. *Ann Intensive Care.* (2018) 8:66. doi: 10.1186/s13613-018-0402-x
25. Shen Y, Huang X, Zhang W. Association between fluid intake and mortality in critically ill patients with negative fluid balance: a retrospective cohort study. *Crit Care.* (2017) 21:104. doi: 10.1186/s13054-017-1692-3
26. Raghu A. Continuous state-space models for optimal sepsis treatment - a deep reinforcement learning approach. *arXiv [Preprint]. arXiv*: 170508422v1. (2017).
27. Joseph Futoma MM. Identifying distinct, effective treatments for acute hypotension with Soda-RL: safely optimized diverse accurate reinforcement learning. *arXiv 200103224.* (2020).
28. Xuefeng Peng YD. Improving sepsis treatment strategies by combining deep and kernel-based reinforcement learning. *arXiv [Preprint]. arXiv*: 190104670. (2019).
29. Zhang Z, Zheng B, Liu N. Individualized fluid administration for critically ill patients with sepsis with an interpretable dynamic treatment regimen model. *Sci Rep.* (2020) 10:17874. doi: 10.1038/s41598-020-74906-z

Conflict of Interest: LS, SL, XZ, LW, WZ, and YLo were employed by Peking Union Medical College Hospital. YLi, MS, and SZ were employed by DHC Software Co., Ltd.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership